

**DISCOVERY AND CHARACTERIZATION OF KNOX PROTEINS
LACKING A HOMEODOMAIN, PRODUCED BY ALTERNATIVE
SPLICING OF KNAT1-LIKE GENES IN GYMNOSPERMS AND
ANGIOSPERMS**

A Dissertation
Presented to
The Academic Faculty

by

Mili Sheth

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
December 2008

**DISCOVERY AND CHARACTERIZATION OF KNOX PROTEINS
LACKING A HOMEODOMAIN, PRODUCED BY ALTERNATIVE
SPLICING OF KNAT1-LIKE GENES IN GYMNOSPERMS AND
ANGIOSPERMS**

Approved by:

Dr. John Cairney, Advisor
School of Biology
Georgia Institute of Technology

Dr. Nael McCarty
School of Medicine
Emory University

Dr. King Jordan
School of Biology
Georgia Institute of Technology

Dr. Jeffrey Dean
School of Forestry and Natural
Resources
University of Georgia, Athens

Dr. Kirill Lobachev
School of Biology
Georgia Institute of Technology

Date Approved: October 27, 2008

ACKNOWLEDGEMENTS

The first person I want to thank is my advisor Dr. John Cairney. He has been a great teacher and mentor and provided me with encouragement and stimulating suggestions throughout the course of this work. He has been constant source of inspiration and has enabled my critical reasoning and thought process.

I would like to thank my committee members Dr. King Jordan, Dr. Kirill Lobachev, Dr. Nael McCarty and Dr. Jeffrey Dean for their invaluable guidance during my thesis work.

I would like to thank my colleague Brande Jones for all the invaluable discussions, support and help during my thesis work. She has been a wonderful friend and always been there to raise my spirits. I would like to thank my former colleagues Barbara Johns, Yiting Zhang and Cuihua Zhu for their help. I would like to thank Dr. Jerry Pullman and Dr. Thomas Oh for their valuable discussions and help throughout my thesis work. I am most grateful to Dr. Jerry Pullman, Georgia Institute of Technology, for providing pine embryos, Dr. Daniel Peterson, Mississippi State University, for loblolly pine genomic DNA, Dr. John Ohlroghee, for *Arabidopsis* RNA from 5 & 10 day old seeds and Dr. Joe Nairn, UGA, for *Arabidopsis* tissue.

I would like to thank my husband, Prarthan Sheth, without whom I would not have been able to reach this stage of my career. I thank him inspiring and supporting me throughout.

I would like to thank my daughter, Riya Sheth, whose innocent smile would wipe away all my frustrations and put things back in perspective. I would also like to thank my family for encouraging me to pursue my dreams and helping me achieve them.

This work has been supported from NSF Grant Award Number: 0543503.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	III
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS	XIII
SUMMARY	XVI
CHAPTER 1 INTRODUCTION	1
1.1 Loblolly pine – the second most important agricultural crop in US	1
1.2 Pine – Arabidopsis Embryos develop similarly but have several distinct features	6
1.3. Plants continue to grow throughout life - Meristems	9
1.4 KNOX genes of homeobox family are important for growth and development	11
1.5 Homeobox family of genes	13
1.5.1 KNOX genes belong to the homeobox family of genes	16
1.5.2 Features of the KNOX family proteins	16
1.6 KNOX genes - Key genes in determining shape and development	19
1.6.1 Classes of KNOX genes – their expression pattern and function	19
1.6.2 Proteins of KNOX family regulate development by interacting with proteins of BELL family and other families	21
1.6.3 Intercellular trafficking of KN1	23
1.6.4 KNOX genes and Hormones	25
1.7 KNOX genes and Evolution	27
1.8 Altered activity of a single gene has profound effects on development	29

CHAPTER 2 KNOX PROTEINS LACKING A HOMEODOMAIN ARE PRODUCED VIA ALTERNATIVE SPLICING OF KNAT-1 LIKE GENES IN GYMNOSPERMS AND ANGIOSPERMS	32
2.1 Introduction	32
2.2 Results	33
2.2.1 PtKN1 – the KNAT1-like mRNA of Loblolly Pine (<i>Pinus taeda</i> L.) Contains Highly Conserved Domains	33
2.2.2 A KNAT1-like mRNA (PtKN1(hd-)) Lacking HD Sequences is Expressed During Pine Embryo Maturation	36
2.2.3 The PtKN1(hd-) mRNA is Expressed Abundantly During Pine Embryo Maturation	39
2.2.4 PtKN1(hd-) Protein is More Abundant than the PtKN1(HD+) Protein in Late Maturing Pine Embryos	39
2.2.5 Genomic Cloning of PtKN1	45
2.2.6 Cloning of an <i>Arabidopsis</i> KNAT1 mRNA lacking HD Encoding Sequences	50
2.2.7 Identification of a Transcript Lacking HD Sequences in a Monocotyledonous Plant Maize RS1	52
2.2.8 Expression of <i>Arabidopsis</i> KNAT1 mRNA lacking HD Sequences	52
2.2.9 The KNAT1 and KNATM Genes in <i>Arabidopsis</i> Exhibit Different Expression Patterns	53
2.2.10 Cloning of <i>Arabidopsis</i> KNAT1 mRNA Lacking the MEINOX encoding sequence	54
2.2.11 Expression of KNAT1 mRNA lacking MEINOX Sequences	54
2.2.12 An OSH1-like Transcript Splice Removes KNOX Sequences and Changes Reading Frame in Rice	63
2.2.13 Proteins Are Encoded by KNAT(hd-) and KNAT(knox-) mRNAs	63
2.3 Discussion	65
2.3.1 KNOX proteins Containing or Lacking HD Sequences are Produced by Alternative RNA Processing of a Single Gene	65

2.3.2 The <i>Arabidopsis</i> KNAT1 Gene Produces mRNA and Protein Lacking HD Sequences	68
2.3.3 In Monocotyledons KNAT1 like Genes are Spliced to Produce Transcripts lacking HDs	69
2.3.4 Splicing Removes MEINOX Sequences from KNAT1 mRNA	70
2.4 Methods and Materials	73
2.4.1 Plants and Embryo tissue	73
2.4.2 RNA isolation and cDNA synthesis	73
2.4.3 Molecular cloning of PtKN1	73
2.4.4 Molecular cloning of KNAT1 spliced forms lacking HD and MEINOX domain encoding sequences	74
2.4.5 Semi-quantitative PCR	75
2.4.6 Western blot analysis	77
2.4.7 Quantification of sq-pcr and western blotting images	78
2.4.8 DNA purification, Cloning and Sequencing	79
2.4.9 Cloning of the Knotted1 gene	79
2.4.10 Alignments	79
2.4.11 Accession Numbers	80
CHAPTER 3 PTKN1 PROMOTER ANALYSIS	81
3.1 Introduction	81
3.2 Materials and Methods	82
3.2.1 Cloning of the Knotted1 promoter	82
3.2.2 5' RACE	83
3.2.3 DNA purification, Cloning and Sequencing	84
3.2.4 KNAT1-like promoter orthologs	84
3.2.5 Comparison of PtKN1 promoter to orthologous KNAT1-like promoters	84

3.2.6 Identification of regulatory motifs from known database	85
3.2.7 Regulatory motif search using known database	85
3.3 Results and Discussion	86
3.3.1 Cloning of PtKN1 gene promoter	86
3.3.2 Identification of transcription start site	88
3.3.3 KNAT1-like Promoter analysis by phylogenetic footprinting	94
3.3.4 In silico PtKN1 promoter analysis - Several cis-regulatory elements are found in KNAT1-like promoters	100
CHAPTER 4 CONCLUSION	110
CHAPTER 5 FUTURE DIRECTIONS	119
APPENDIX A SUPPLEMENTARY TABLE AND FIGURES FOR CHAPTER 2	127
Figure A.1: Nucleotide and deduced amino acid sequence of PtKN1(HD+).	129
Figure A.2: Nucleotide and deduced amino acid sequence of PtKN1(hd-).	131
Figure A.3: Alignment of the nucleotide sequences of PtKN1(HD+) and PtKN1(hd-).	133
Figure A.4: Peptide sequence used for making antibody for PtKN1(HD+) and PtKN1(hd-) proteins.	134
Figure A.5: Genomic sequence map.	135
Figure A.6: Alignment of the nucleotide sequences of <i>Arabidopsis</i> KNAT1(hd-) and KNAT1 full length mRNA.	136
Figure A.7: Alignment of the nucleotide sequences of <i>Arabidopsis</i> KNAT1(hd-), KNAT1 full length mRNA and KNAT1 gene. (page 1)	137
Figure A.8: Alignment of the amino acid sequences of maize RS1 (NP_001105331.1) and deduced amino acid sequence of HD splice form mRNA (DY236300.1).	140
Figure A.9: Alignment of the nucleotide sequences of maize RS1(hd-) (DY23600.1) and RS1 full length mRNA (NM_0011118.1).	141
Figure A.10: Alignment of the amino acid sequences deduced from KNAT1(knox-) and <i>Arabidopsis</i> KNAT1 (NM_116884).	142

Figure A.11: Alignment of the nucleotide sequences of <i>Arabidopsis</i> KNAT1(knox-) and KNAT1 full length mRNA.	143
Figure A.12: Alignment of the nucleotide sequences of rice OSH(knox-) (AK107637.1, NM_001057674) and OSH full length mRNA (D16507.1).	144
Figure A.13 Amino acid alignment of KNAT1 with KNAT2 and KNAT6 in the region of epitope binding.	145
Figure A.14 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles.	146
Figure A.15 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles.	147
Figure A.16 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles.	148
APPENDIX B Genomic SEQUENCES	149
APPENDIX C PROTOCOLS	152
RNA isolation by TRI reagent	152
SMART RACE	152
DNA and gel purification	153
Cloning	154
Colony PCR	154
Plasmid preparation and purification	154
Protein isolation by TRI reagent	155
Protein Concentration Assay	156
Western Blotting	156
Preparing genomic libraries for genomic walk	158
REFERENCES	160

LIST OF TABLES

Table 3.1: Conserved regions found by footprinter in KNAT1-like promoters could be potential regulatory motifs.	98
Table 3.2: Regulatory motifs found in PtKN1 promoter using PLACE database.	101
Table 3.3: Regulatory motifs found in PtKN1 promoter using PLANTCARE database.	103
Table 3.4: Regulatory motifs found in KNAT1-like promoters.	105

LIST OF FIGURES

Figure 1.1: Loblolly pine grows over wide geographic range [3].	1
Figure 1.2: Somatic embryogenesis [11].	3
Figure 1.3: Comparison of zygotic and somatic embryo development stages in loblolly pine [11].	5
Figure 1.4: Embryogenesis - gymnosperm vs. angiosperm [30].	8
Figure 1.5: Plant growth proceeds through regions of active growth – Meristem [52].	10
Figure 1.6: Mutation in KNOX genes affects many aspects of plant development [50, 69].	12
Figure 1.7: KNOTTED1 belongs to the KNOX family of genes.	15
Figure 1.8: Features of knox gene family.	18
Figure 1.9: HD proteins bind DNA as dimer [116].	18
Figure 1.10: Altered activity of a single gene has profound effects on development [188].	30
Figure 2.1: Alignment of the amino acid sequences deduced from the sequences of PtKN1 and <i>Arabidopsis</i> KNAT1 (NM_116884) mRNA.	35
Figure 2.2: Alignment of the amino acid sequences deduced from the sequences of PtKN1(HD+) and PtKN1(hd-) mRNA.	37
Figure 2.3: Domain structure of PtKN1(HD+) and PtKN1(hd-) mRNA.	38
Figure 2.4: PtKN1(HD+) and PtKN1(hd-) mRNAs and proteins are developmentally regulated.	44
Figure 2.5: A single PtKN1 gene produces two mRNAs – PtKN1(HD+) and PtKN1(hd-).	47
Figure 2.6: Restriction digestion of different regions of genomic DNA to determine PtKN1 structure.	49
Figure 2.7: Alignment of the amino acid sequences deduced from <i>Arabidopsis</i> KNAT1(hd-) and KNAT1 (full length) cDNA.	51

Figure 2.8: Spliced forms in <i>Arabidopsis</i> KNAT1 that lack HD or KNOX regions.	62
Figure 3.1: Promoter isolation by genomic walk.	87
Figure 3.2: Determine the transcription start site of PtKN1 transcripts.	89
Figure 3.3 Transcription start sites in the promoter region and protein binding sites and stem-loop structures possibly involved in regulation.	94
Figure 3.4: Promoter analysis by phylogenetic footprinting using footprinter program.	97
Figure 3.5: Regulatory motif search in KNAT1-like promoters from different species.	105

LIST OF ABBREVIATIONS

aa	Amino acid
ABA	Abscisic acid
ABRE	Abscisic acid Response Element
ARI	ARIADNE gene
AtOFP	<i>Arabidopsis thaliana</i> Ovate Family Protein
BCP	1- Bromo-3-Chloropropane
BLAST	Basic Local Alignment Search Tool
BLR	Bellringer
bp	Base pair
BP	Brevipedicellus
Cdc2	Cell Division Cycle2
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
GA	Gibberellic acid
GA20ox	GA-20 oxidase
GARE	Gibberellin Response Element
GUS	uidA gene encoding the b-glucuronidase enzyme
EtBr	Ethidium Bromide
HBK	Homeobox of KNOX class
HD	Homeodomain
hd-	HD lacking
HD+	HD containing
HD-Zip	HD Zipper
IPT	Isopentyl transferase
IRO	Iroquis
KN1	Knotted1

KNAT	Knotted-like in <i>Arabidopsis thaliana</i>
KNOX	Knotted-like homeobox
LeT6	<i>Lycopersicon esculentum</i> T6
MEIS	Myeloid Ecotropic Integrase Site
mRNA	messenger RNA
MYA	Million Years Ago
MYB	Myeloblastosis
MYC	Myelocytomatosis
NLS	Nuclear Localization Signal
nt	Nucleotide
NTH	<i>Nicotiana tobaccum</i> Homeobox
NUP	Nested Universal Primer
ORF	Open Reading Frame
osh	<i>Oryza Sativa</i> Homeobox
PAGE	Polyacrylamide Gel Electrophoresis
PBC	Pre-B Cell
PCR	Polymerase Chain Reaction
PHD	Plant HD
PNF	Poundfoolish
PNY	Pennywise
POTH1	Potato Homeobox1
PtKN1	<i>Pinus taeda</i> Knotted1
PTS	Petroselinum
RACE	Rapid Amplification of cDNA Ends
RAM	Root Apical Meristem
RNA	Ribonucleic Acid
rs1	rough sheath1
RT-PCR	Reverse transcription PCR
SAM	Shoot Apical Meristem
SDS	Sodium Dodecyl Sulphate
SMART	Switching Mechanism At 5' End of RNA Transcript

SORLIP1AT	Sequence Over-represented in Light-Induced Promoters
STM	Shoot Meristemless
TALE	Three Amino acid Loop Extension
TBS	Tris-Buffered Saline
TBS-T	TBS – Tween
TPP	Thiamin pyrophosphate
TSS	Transcription start site
UPM	Universal Primer Mix
UTR	Untranslated region
VAN	Vamana
Zm	Zea mays

SUMMARY

Embryogenesis in Gymnosperms and Angiosperms while broadly similar differs in important respects. Despite these differences, the genes expressed in developing pine embryos show strong sequence similarity to those found in angiosperm genomes. The distinguishing features of conifer embryogenesis may be a consequence of the level, timing or localization of expression of certain regulatory genes perhaps supplemented by the activity of a few unique or highly differentiated proteins.

Homeobox genes encode homeodomain (HD) proteins which function as transcription factors and play an important role in plant and animal development by controlling cell specification and pattern formation. The KNOX (knotted-like homeobox) family of regulatory genes belongs to a homeobox gene family which exerts considerable influence upon plant development. Alteration in the expression of KNOX genes, through mutation or by increasing or decreasing mRNA levels in transgenic plants, produces profound changes in morphology and growth. The ability of these proteins to affect a broad range of cellular activities is explained in great measure by the presence within the KNOX proteins of distinct domains that can interact with a variety of other regulatory proteins. The HD region of KNOX proteins is a sequence-specific DNA-binding domain that can directly regulate expression of target genes. The KNOX1 and KNOX2 regions constitute the MEIKNOX domain which has been shown to homodimerize with KNOX family proteins or heterodimerize with members of the BELL family of TALE homeodomain proteins.

I have cloned (Knotted1 in *Arabidopsis thaliana*) KNAT1-like mRNAs which lack HD sequences from embryos of loblolly pine (*Pinus taeda* L.), hereafter referred as PtKN1(hd-). Production of PtKN1(hd-) mRNAs is developmentally regulated and their encoded protein is abundant in mature pine embryos. The PtKN1 gene has 5 exons; KNOX1 and KNOX2 domains are encoded by the first three exons, and the large exon 3 separates these from exons 4 and 5 which encode HD sequences. The PtKN1(hd-) mRNA has a unique 3'UTR which derives from the proximal region of 'exon 3'. Both forms of PtKN1 are produced by the same gene; the regulatory dynamic is between cleavage-polyadenylation within intron 3 to produce PtKN1 mRNA lacking HD sequences and splicing of exon 3 to exon 4 which excludes the 3'UTR/exon3 sequence to create an mRNA which encodes a HD. I have identified and characterized a KNAT1 mRNA in *Arabidopsis* which lacks HD sequences. While KNAT1 has been studied for many years, this is the first report of a KNAT1 mRNA lacking HD. The expression pattern of AtKNAT1(hd-) mRNA differs from that of the KNATM-B mRNA, which encodes a KNOX gene lacking HD sequences, implying a distinct role for these two genes. While KNATM appears unique to dicotyledons, I identified a KNAT1 mRNA lacking HD sequences for the RS1 gene of maize, a monocotyledon. This is the first report of splicing of KNAT1 genes to produce mRNAs lacking HD sequences. The phenomenon appears to be ubiquitous as it is observed in gymnosperms, and both dicotyledonous and monocotyledonous angiosperms.

CHAPTER 1

INTRODUCTION

1.1 Loblolly pine – the second most important agricultural crop in US

Loblolly pine is one of the major timber species in the US with 1.5 billion seedlings planted each year. It provides 1.7 million jobs with the annual payroll of \$51 billion. Pine is the 2nd most important agricultural crop in US after corn and grows over a wide range from New Jersey to Texas (Figure 1.1) [1, 2].

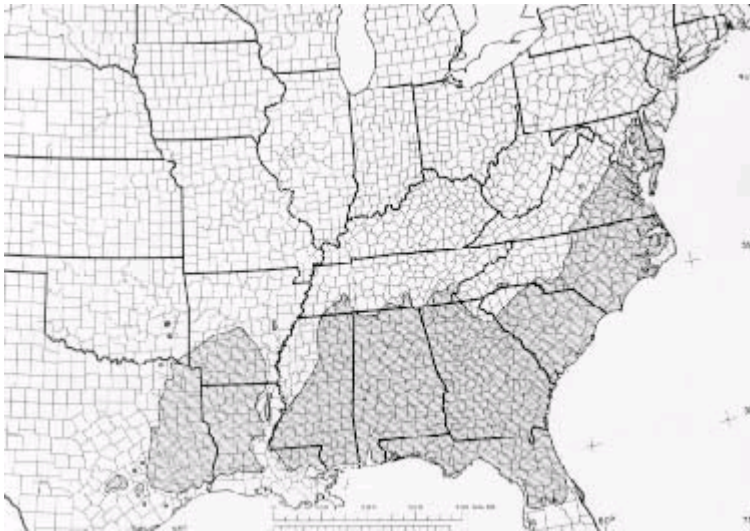


Figure 1.1: Loblolly pine grows over wide geographic range [3].

Loblolly pine grows over a wide range from southern New Jersey to central Florida and eastern Texas (shown as shaded region in the map above).

Pine wood is very useful for pulp, paper, furniture, plywood, poles, boxes, pallets etc. In addition pine provides water quality improvement, land reclamation and rehabilitation, agroforestry, aesthetics, recreational opportunities, place for wildlife as well as food for the cattle [1, 4]. Recently the use of pine as a source of biomass for biomaterial and bioenergy is also being investigated [5]. The increasing demand of forest products and decreasing availability of land-base for producing trees will cause difficulties for the forest product industries. Therefore need arises to propagate large number of pine trees and clonal propagation techniques are expected to play a significant role in providing genetically superior pine trees [6, 7]. Clonal plantation can be established by organogenesis and embryogenesis. In 1960s, the first plantlets from forest tree tissue culture were produced by organogenesis using adventitious shoots. Organogenesis includes outgrowth and division of seedling shoot, using callus and cell cultures to regenerate adventitious shoots and complete plants, using cultured cotyledons and entire embryos to induce adventitious buds and using cultured shoot tips to induce adventitious and axillary buds. Stem cuttings, leaf cuttings, root cuttings, root suckers, root-stem cuttings, grafting and air-layering are some other methods of clonal propagation [8]. The most common method used is somatic embryogenesis in which embryo formation in seeds can be mimicked in lab and was first developed for carrot propagation [9]. Somatic embryogenesis for Loblolly pine was first reported 20 years ago [10].

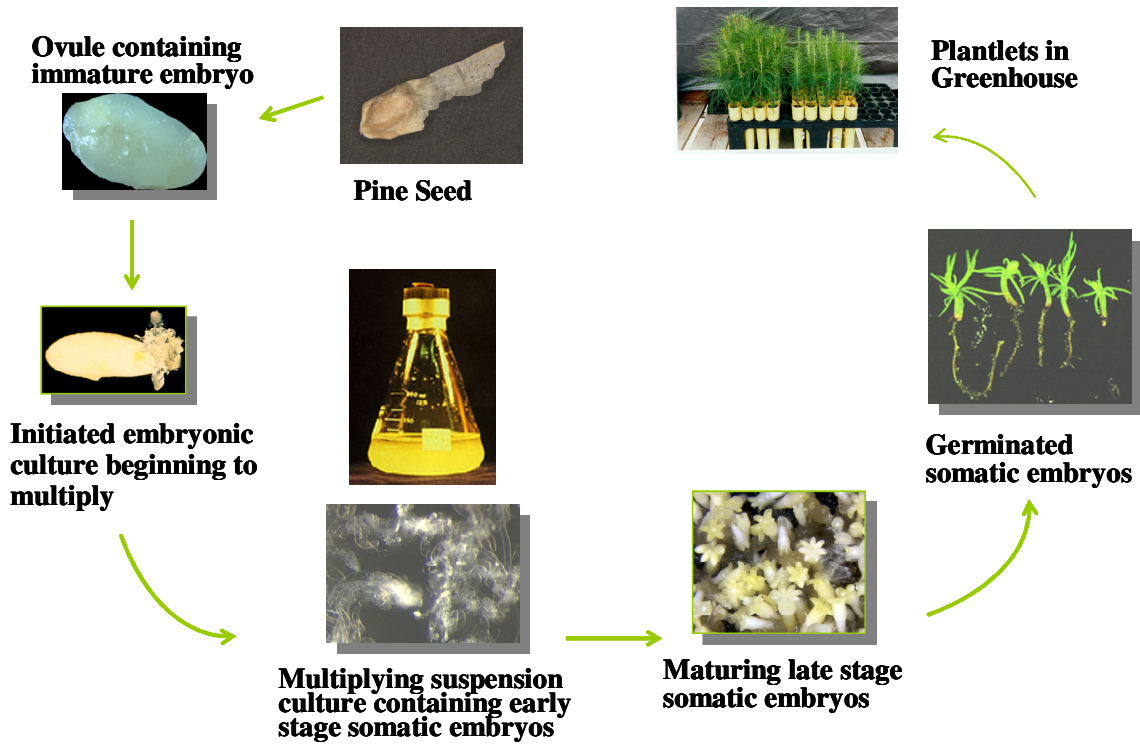


Figure 1.2: Somatic embryogenesis [11].

In this process, pine seeds are extracted from pine cone and cracked to reveal the ovule containing embryo. The embryos are then placed on initiation medium where they begin to multiply. Embryos in early stages are transferred to liquid multiplication medium and later to solid maturation medium. Once they germinate the resulting plantlets are transferred into soil.

In this process as shown in Figure 1.2 pine seeds are extracted from pine cone and cracked to reveal the ovule containing an embryo which is then placed onto an initiation medium. Later, the embryo is transferred to multiplication and maturation medium and after germination the resulting plantlets are transferred into soil. The process is rapid and has potential to provide thousands of high quality embryos which are genetically identical. The process has potential to provide genetically improved trees. However for reasons unknown there are some drawbacks in this technique which limits the commercial production of pine trees. Pine zygotic embryo development has been staged from 1 to 9.12 as shown in Figure 1.3. Stages 9.1 to 9.12 are morphologically similar but biochemically different. The zygotic and somatic embryo developmental stages for loblolly pine are comparable and as shown in Figure 1.3 (Pullman and Webb 1994). The success of somatic embryogenesis is very dependent on genotype [12]. The somatic embryos do not mature fully, even the best performing genotypes only reach stage 9.1 and the quality of embryos can be poor. Other limiting factors are low initiation rate, low culture survival, low maturation rate, low germination, low or no embryo production, [11, 13-16]. Study of the molecular biology of embryogenesis will provide fundamental knowledge which will assist the improvement of tree propagation strategies. Molecular biology studies of the zygotic embryogenesis will be used as model to understand the somatic embryogenesis [17, 18].

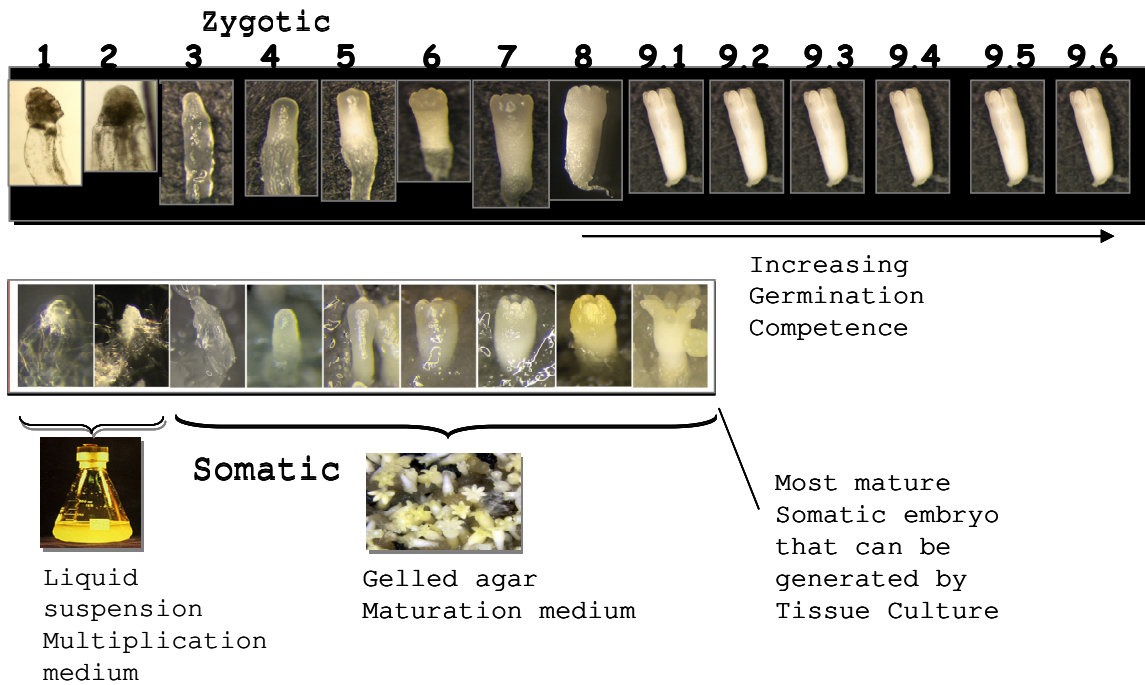


Figure 1.3: Comparison of zygotic and somatic embryo development stages in loblolly pine [11].

The zygotic embryos develop from stage 1 to 9.12. Stage 9.1 onwards the embryos are morphologically similar. Somatic embryos develop similarly but stop at stage 9.1.

1.2 Pine – Arabidopsis Embryos develop similarly but have several distinct features

Embryogenesis, the process of growth and differentiation is an important period of plant development because during this phase the meristem and shoot-root body pattern of a plant are specified. Traditionally plant embryogenesis is perceived as being composed of two phases: a morphogenic phase and a maturation phase. The overall body plan is established during the early morphogenic phase i.e. division of 3 primary tissues: the outer protoderm, the inner ground tissue and the central procambium (which forms the epidermis, the cortex and the endoderm, and the vascular tissue, respectively) [19]. Metabolic changes such as storage protein accumulation and acquisition of desiccation tolerance take place during the maturation phase [20]. After maturation the embryo undergoes desiccation and dormancy for a period of time until conditions are favorable [21]. Based on the differences in process of embryogenesis the plant kingdom is divided into angiosperms and gymnosperms.

Embryo developments in angiosperms and gymnosperms have several unique features. Angiosperms undergo double fertilization and produce a diploid embryo by fertilization of one sperm with the haploid egg cell and a triploid endosperm (nutritive tissue surrounding the embryo) by fertilization of the second sperm with the two polar nuclei. Angiosperm seeds are enclosed in ovary of the mother plant [22-24]. Gymnosperms undergo single fertilization to produce a diploid embryo and haploid female gametophyte (endosperm equivalent nutritive tissue surrounding the embryo). Gymnosperms have naked seeds that develop outside the ovary [25]. The first asymmetric division of angiosperm produces smaller apical cell and larger basal cell which gives rise to embryo proper and the suspensor tissue plus some root tissue

respectively [26]. Gymnosperms undergo several nuclear divisions without cytokinesis to enter a free nuclear phase after fertilization, followed by cellularization to form eight-celled proembryo arranged in two tiers which undergoes division to form four tiers. The four cells towards the micropylar end give rise to embryo proper and the next layer of cells produce the suspensor. The period of embryogenesis from fertilization to suspensor elongation is called proembryony and from the suspensor elongation to root meristem establishment is the early embryogeny [25, 27, 28]. Formation of multiple embryos is common for gymnosperms but uncommon in most angiosperms. Eventually only one embryo is produced per seed and the others die by programmed cell death [29]. Polyembryony occurs by fertilization of the egg cells within one or more archegonia by different pollen grains to produce zygotes that have different genotypes within the same seed or by multiplication of immature embryos. The former is known as simple embryony and the latter is cleavage polyembryony [27]. Angiosperm and gymnosperm embryos also have morphological differences. *Arabidopsis* has a conspicuous embryo with heart, torpedo and bent cotyledonary stage while pine embryos are multicotyledonary and have 6 to 8 cotyledons which enclose the SAM [19, 25]. Mature embryo is a multicotyledonary embryo with cortex, epidermis, root cap, shoot and root meristem and hypocotyl. Cellular differentiation in embryonic mass and development of embryo in preparation for seed dormancy occurs during the late embryogeny [25]. Comparison of angiosperm and gymnosperm embryogenesis is shown in Figure 1.4.

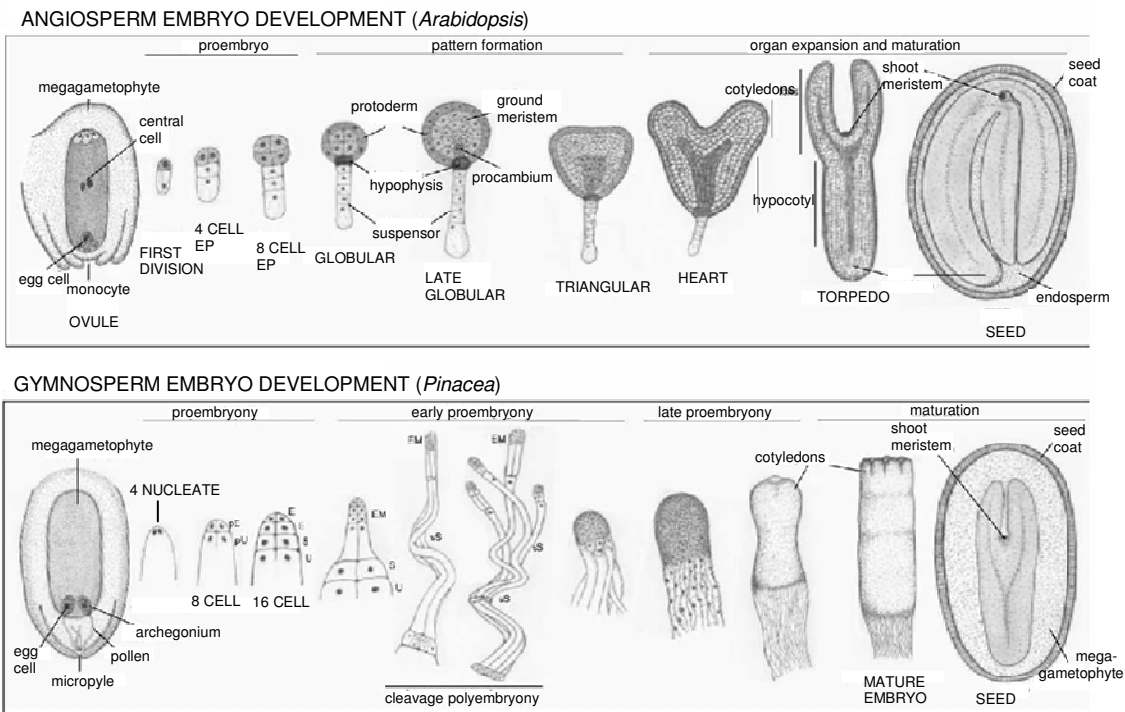


Figure 1.4: Embryogenesis - gymnosperm vs. angiosperm [30].

The embryogenesis process in gymnosperms and angiosperms are similar but have several distinct features. While gymnosperms undergo single fertilization to produce diploid embryo and haploid female megagametophyte (FMg), the angiosperms undergo double fertilization to produce a diploid embryo and triploid endosperm. In gymnosperms FMg develop before fertilization whereas in angiosperms endosperm development after fertilization; initially polyembryony (1-4) vs. single embryo; Morphological differences can be observed in the embryos in the later stages. 6-8 cotyledonary primordia grow to form cotyledons and enclose the SAM in gymnosperm vs. conspicuous embryo with heart, torpedo and bent cotyledonary stage (dicot) are seen in angiosperms.

Despite the differences in pine and *Arabidopsis* embryogenesis, the genes expressed in developing pine embryos show strong sequence similarity to those found in angiosperm genomes [31]. These differences between pine and *Arabidopsis* embryogenesis can be explained by differences in the level, timing or localization of expression of certain regulatory genes or through the agency of a few unique or highly differentiated proteins. The homeobox genes are a family of regulatory genes which are found in animals, plants and fungi and exert considerable influence upon their development [32-39].

1.3. Plants continue to grow throughout life - Meristems

The body plan of most animals i.e. formation of organs like head, hands, legs, or wings can be recognized during embryogenesis. Growth, differentiation and maturation leads to further development of the adult organism [40]. Higher plants continually develop throughout their life and repeatedly produce new structures. These structures are formed by indeterminate and self-renewing groups of cells called meristem. Vertical growth proceeds from apical meristems that are located at shoot and root apices and are called shoot apical meristem (SAM) and root apical meristems (RAM) respectively [41-44] , see Figure 1.5. A different type of meristem is cambial meristem which forms a cylinder around the stem and controls production of xylem and phloem cells [45]. Shoot apical meristem produces the above ground parts of plants like stem, leaves and flowers. In order to understand organogenesis events that determine the overall development of plant, careful examination of the events occurring in the meristem is required. Complex gene expression patterns regulate the formation of SAM and maintain the population of indeterminate cells. Failure to form and maintain SAM results in abnormal organization

of shoot structure [46]. Therefore it is important to study the genes that regulate the SAM formation and maintenance. Maize gene *kn1* which belongs to the KNOX family of homeobox gene family has been well characterized and has been shown as a useful molecular marker for maize SAM [47-49]. Mutation or overexpression of KNOX genes in transgenic plants causes great difference in size, shape and development [50, 51].

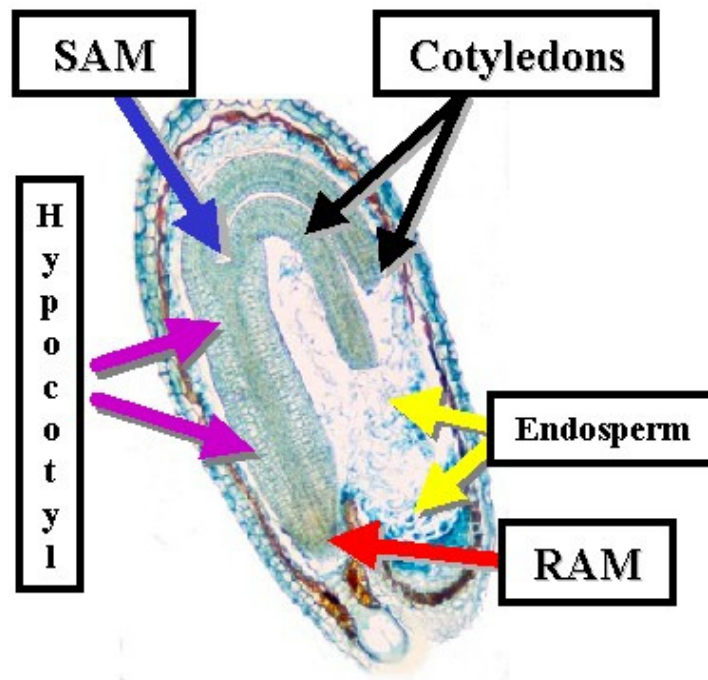


Figure 1.5: Plant growth proceeds through regions of active growth – Meristem [52]. Meristems are the region of active growth in plants. Vertical growth proceeds from SAM and RAM the positions of which are shown in an angiosperm embryo here. Cambial meristem forms the xylem and phloem.

1.4 KNOX genes of homeobox family are important for growth and development

Various mutational studies have shown that plant KNOX genes play an important role in regulating meristem function. Their loss or gain of function mutation can affect overall plant height, leaf shape, meristem development and floral development as shown in Figure 1.6. Loss-of-function mutants of maize *kn1* lead to formation of fewer lateral meristem but more lateral organs like leaves and carpel [53]. Their gain-of-function mutants lead to formation of ectopic meristems, inhibition of leaf differentiation and knot-like meristematic structure formation in leaf veins [48, 54]. Strong loss-of-function mutants of *STM* show deficiencies in SAM development and result in severely fused cotyledonary petioles [55, 56]. Less severe mutations result in a disorganized SAM from which ectopic lateral organ (leaves during vegetative growth, flowers during reproductive growth) formation takes place [57, 58]. Gain-of-function mutation of *STM*, *KNAT1*, *KNAT2* or *KNAT6* results in small malformed leaves with extensive lobing around the leaf margin due to inhibition of proper leaf cell expansion and differentiation [44, 59-63]. Loss-of function mutation of *KNAT1* result in reduced internode length and bends in stem tissue and are overall significantly shorter than wild-type plants [64, 65]. Overexpression of several KNOX genes from *Arabidopsis*, maize, rice and tobacco induce ectopic SAM formation on adaxial leaf surfaces [44, 66-68].

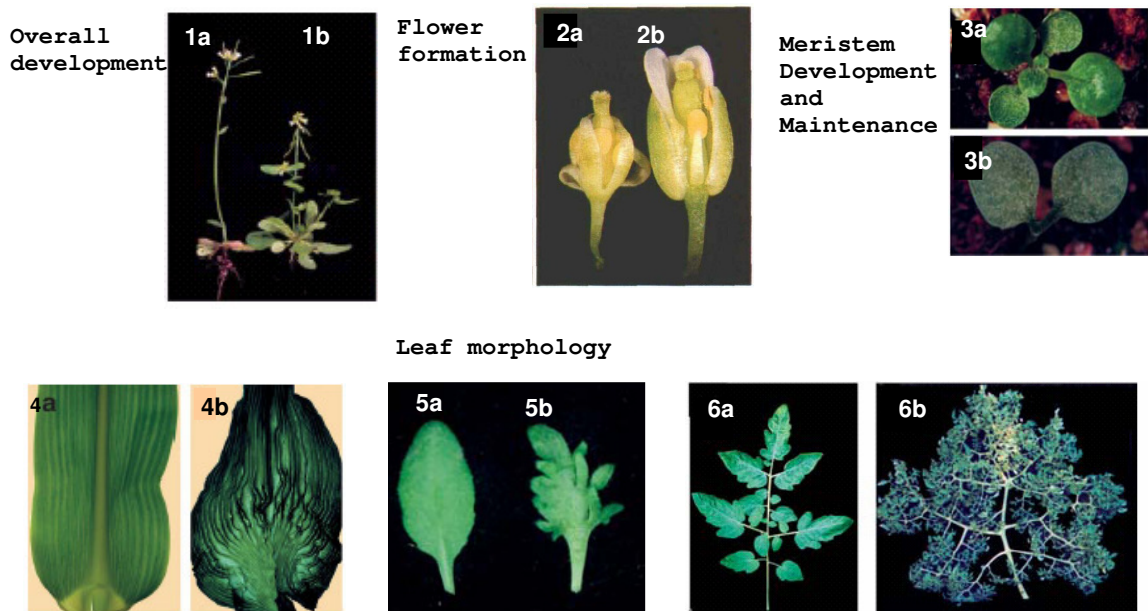


Figure 1.6: Mutation in KNOX genes affects many aspects of plant development [50, 69].

1a - Wild-type *Arabidopsis* plant

1b - *Arabidopsis* bp mutant showing dwarf phenotype with shorter and downwards pointing pedicels and shorter internodes

2a - *Arabidopsis* KNAT1 with severely lobed transformant flower

2b - Wild-type *Arabidopsis* flower

3a - Top view of normal *Arabidopsis* seedling

3b - *stm-1* mutant with no SAM and fused cotyledons

4a - Wild-type maize leaf blade

4b - KN1-N2 leaf blade showing knotted phenotype

5a - Wild type *Arabidopsis* leaf

5b - 35S:BP leaf shows lobed phenotype due to misexpression of BP

6a - Wild-type tomato leaf has 8-16 leaflets

6b - 35S:KN1 tomato leaf is severely dissected and has upto 1000 leaflets

For gymnosperms, the functions of HBK1 and HBK3 from Norway spruce have been shown by transformation in *Arabidopsis*. HBK1 when ectopically expressed in *Arabidopsis* show defects in leaf development [70]. HBK3 when overexpressed in *Arabidopsis*, plants were shorter with reduced apical dominance, lobed leaves and short petioles. Elevated expression of HBK3 protein in somatic embryos of Norway spruce promoted differentiation of immature embryos increased the frequency with which mature embryos were observed and produced an enlarged stem apical meristem [71]. In cell lines expressing antisense HBK3 constructs (HBK3-A), the differentiation of immature embryos was greatly reduced [71]. Norway spruce HBK2 plays a role in somatic embryo development [72].

Above suggests that knox genes play a significant role in overall development, meristem formation and maintenance, flower and leaf development and morphology and embryo development. Alteration in the expression of kn1 genes through mutation or by overexpression in transgenic plants produces profound changes in morphology and growth. Therefore kn1 is an important gene and could provide clues to embryo development in gymnosperms and angiosperms.

1.5 Homeobox family of genes

The homeobox is a 180 bp consensus DNA sequence present in several genes involved in developmental processes [73, 74]. Homeobox genes were first identified in *Drosophila melanogaster* to cause homeotic transformations of the body segment [75]. Homeobox genes encode HD proteins which function as transcription factors and play an important role in plant, fungi and animal development by controlling cell specification and pattern formation [32-39]. Transcription factors are proteins that bind to promoters of

target genes in a sequence specific manner and either activate or suppress their expression [76]. The homeobox encodes a 60 amino acid motif called the HD which is arranged in three α -helical structure separated by a loop and a turn. Specific sequence in the third helix of the HD plays a significant role in sequence-specific DNA recognition and binding [73]. The HD is a sequence specific DNA-binding domain and can directly regulate expression of target genes [77, 78].

The HD family is divided into two classes, the typical HD class and the TALE (three amino acid loop extension) class which is the atypical HD class. The TALE class is characterized by three extra amino acids PYP (proline-tyrosine-proline) between helix 1 and 2 of the HD. Based on the HD sequence, several classes have been identified for TALE superfamily: PBC (Pre-B cell) [37], MEIS (Myeloid ecotropic viral integration site) [36], TGIF [32], IRO (Iroquis) [38], MOHAWK (MKX) [39] in animals; the mating type genes (M-ATYP) [33] and CUP genes in fungi; and KNOX [34] and BELL [35] in plants. Examples of the TALE family which play significant role in cell fate specification are human PRL gene and the *Saccharomyces pombe* MATi gene [79-81]. The typical HD proteins, unlike the atypical HD proteins, do not have these extra amino acids. An example of one such typical HD protein is Antennapedia from *Drosophila* which controls the placement of legs [75]. Typical HD families in plants include the PHD finger families (plant HD finger proteins) which have the zinc-finger motif [82, 83]; the HD-Zip family (HD zipper proteins) which have the leucine zipper [84-87] and the *Glabra2* family of homeobox genes which have leucine zipper like motifs [88]. In plants the TALE superfamily is divided into KNOX and BELL classes of transcription factors. The KNOX class is further divided into class I and II based on their sequence similarity, expression

pattern and position of intron [34, 89]. **KNOTTED1** belongs to class I **KNOX** genes. See Figure 1.7.

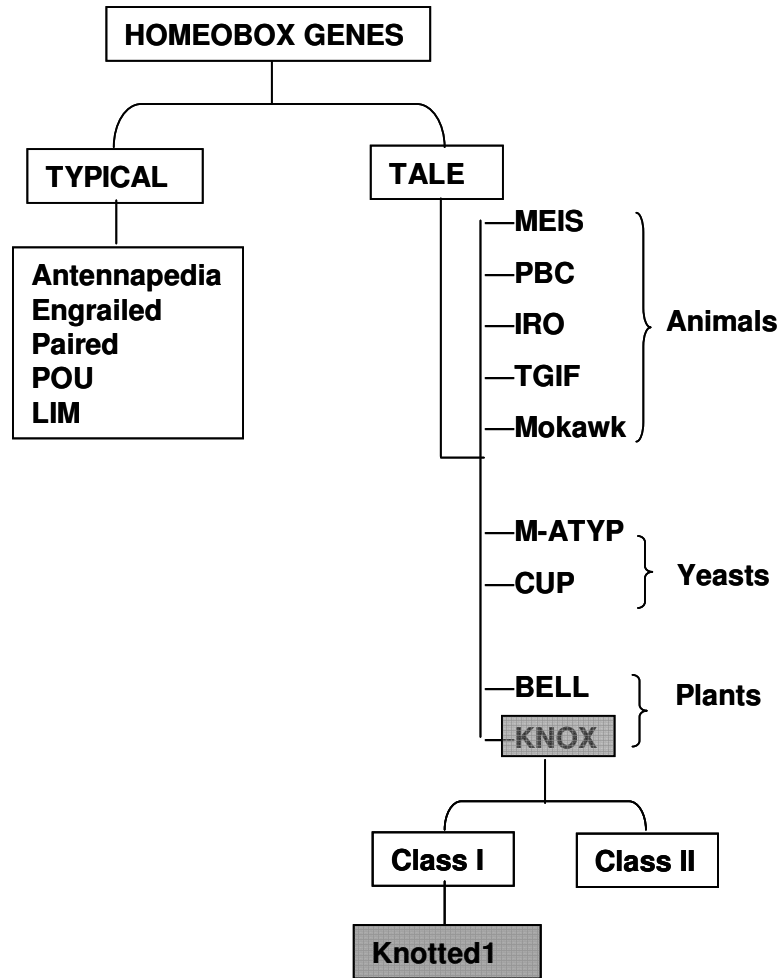


Figure 1.7: KNOTTED1 belongs to the KNOX family of genes.

Homeobox genes are divided into two classes: typical and TALE family. The TALE family is very diverse and has several members in plant, animals and yeast kingdom. **KNOX** genes belong to TALE family of homeobox genes and **KNOTTED1** is a **KNOX** gene.

1.5.1 KNOX genes belong to the homeobox family of genes

Homeobox genes were first identified in *Drosophila* to cause homeotic transformations of the body segment [75]. Homeobox genes encode HD proteins that function as transcription factors and play an important role in plant and animal development by controlling cell specification and pattern formation. A large group of transcription factors called KNOX genes that are found in plants belong to the homeobox gene family. The presence of HD proteins in plants was first recognized in maize after the cloning of Knotted1 (Kn1) by transposon tagging [90]. Kn1 belongs to the (KNOX) family proteins. Several KNOX genes have been isolated from various plants such rice [91], *Arabidopsis*[60], soyabean [92], barley [93], tomato [94], tobacco [95], spruce (a gymnosperm) [70, 72] and various other gymnosperms [96].

1.5.2 Features of the KNOX family proteins

The proteins of KNOX family have several conserved regions such as a MEIKNOX domain, an ELK domain HD, (See Figure 1.8). The MEIKNOX domain can be divided into two conserved region KNOX1 and KNOX2 separated by a less conserved spacer [97]. The KNOX1 domain is thought to be involved in suppression of its target gene [98] and the KNOX2 domain is thought to be involved in dimerization and transactivation [99]. The plant KNOX proteins and human MEIS (myeloid ecotropic viral integration site) proteins are closely related due to their conserved N-terminus region, the KNOX and MEIS domains respectively [97]. The MEIKNOX domain has been shown to homodimerize with KNOX family proteins [93, 98] or heterodimerize with BELL family of TALE HD proteins [93, 100-105]. The ELK domain based on its predicted secondary structure might be involved in protein-protein interaction [34, 98, 99, 106] or it may act

as transcriptional repressor [98]. Some authors have proposed that the ELK domain has the nuclear localization signal [107]. Sakamoto et al. (1999) generated chimeric proteins by exchanging different regions of KNOTTED1-like HD proteins in tobacco and they observed that the abnormal phenotypes observed were most severe when C-terminal half of ELK domain was exchanged.

The HD is a three- α helical structure and the third helix functions as the DNA-recognition helix [73, 102] (see Figure 1.9). Most TALE HD proteins have the conserved WFXN sequence in the third helix, where X is the 50th amino acid. In many cases this position is occupied by a small, non-polar residue and is critical for DNA-binding specificity of the HD [97, 108]. Almost all HD proteins have proline-tyrosine-proline conserved in positions 24 to 26. Amino acids WW (helix 1), PYP (loop), WFIN (helix 3) are important for DNA binding; and in particular PYP and WF are conserved among TALE family proteins in plants and animals [109]. In the animal HD proteins it has been suggested that the N-terminal end of the HD is required for recognition of the target sequences [74, 109, 110]. It has also been suggested that the basic amino acid cluster in the N-terminal HD region functions as a nuclear localization signal [111]. In animals, the MEIS HD proteins have isoleucine at the 50th position that specifically recognizes the TGACAGG/CT motif, and the PBX HD proteins have glycine that specifically recognizes TGAT motif [112]. KN1 from maize and STM and KNAT1 from *Arabidopsis* have been shown to interact with TGACAGG/CT motif [102, 103], suggesting that MEIS and KNOX recognize the same DNA motif. OSH1 from rice has been shown to recognize a DNA motif similar to that recognized by MEIS/KNOX [98] and NTH15 from tobacco has been shown to interact with TGAC sequence. MEINOX domain by

itself can form heterodimers but needs the HD for homodimerization [98, 100, 102, 113, 114]. The spliced form of human Meis2 gene which has MEINOX domain but has partial HD deletion is unable to bind DNA and by competing with the HD containing form for its interaction partners, functions as a dominant-negative regulator [115]. At the start of this project no splice which removed HD sequences had been identified for any plant knox gene.



Figure 1.8: Features of knox gene family.

The proteins in KNOX family possess MEINOX domain which is divided into KNOX1 and KNOX2 regions separated by a linker which are involved in suppression of target gene and dimerization, ELK domain which is involved in protein-protein interaction and the HD which is the DNA binding domain.

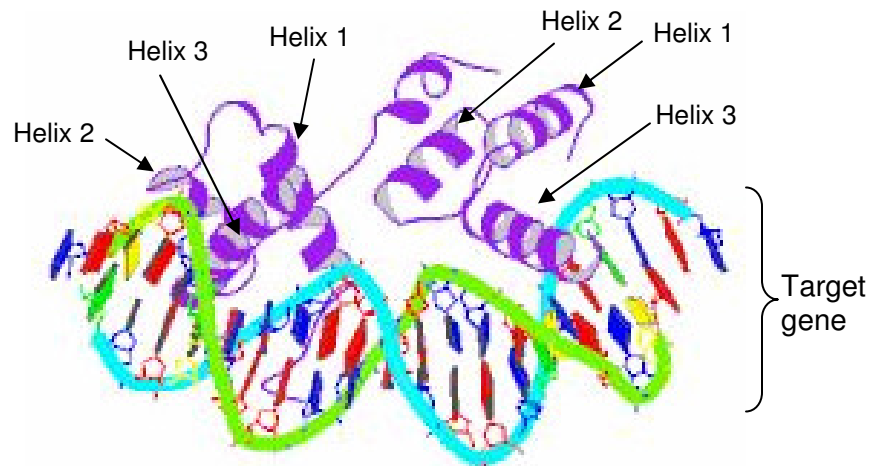


Figure 1.9: HD proteins bind DNA as dimer [116].

The three helices in purple connected by loop represent the HD. HD proteins form homo or heterodimers and as dimer bind to the DNA to activate or suppress the expression of target gene. The binding occurs through the third helix of the HD.

1.6 KNOX genes - Key genes in determining shape and development

1.6.1 Classes of KNOX genes – their expression pattern and function

KNOX genes are divided into two classes - class I and II; based on their sequence similarity, expression pattern and position of intron [34, 89]. Class I KNOX genes are expressed in the shoot apical meristem; and are required for their development and maintenance [34, 117]. *Arabidopsis* has four class I KNOX genes - STM (shootmeristemless) [55, 56], KNAT1 (Knotted-1 like in *Arabidopsis thaliana*), KNAT2 [60] and KNAT6 [61] which have minor sequence differences. Class I KNOX genes from monocot and dicot species have specific or redundant function due to their distinct and overlapping expression patterns. During *Arabidopsis* embryogenesis, STM is first activated at the late globular stage in the center of the apical domain that will give rise to SAM [46]. KNAT1 is first activated and shows highest expression in late globular to early heart stage and thereafter gradually decreases [118]. Neither STM nor KNAT1 are expressed in developing cotyledons during *Arabidopsis* embryogenesis [46]. During vegetative growth, STM is expressed in the central and peripheral zone throughout the SAM but absent in leaves and the meristem cells destined to become the next leaf [48, 56, 60]. KNAT1 is expressed in the peripheral and rib zone of SAM, and immediately below SAM [60]; KNAT2 is expressed in SAM [63, 119] and KNAT6 in shoot apex, root and various organs and tissues [61]. During reproductive growth, STM is expressed in inflorescence and floral meristems, whereas KNAT1 is expressed in a narrow region of cells in the inflorescence meristem but not in the floral meristem [56, 60]. Thus the class I KNOX genes in *Arabidopsis* show similar but slightly different pattern of expression in SAM. Discrete patterns of KNOX expression have also been seen in maize. ZmKN1 and

ZmKNOX8 are expressed in SAM and developing stem but are absent from leaf primordial and mature leaves. ZmRS1 and ZmKNOX3 are also spread out in SAM in areas that overlap with ZmKN1. In addition they are also expressed at the base of leaves, internodes and developing axillary shoot meristems [34, 47]. Class I KNOX genes in tobacco such as NTH1 and NTH15 in the central and peripheral zone of SAM but are absent in the leaf founder cells. NTH19 is expressed in the rib zone which contributes to the internal stem tissue and NTH20 is expressed in the peripheral zone of SAM from which organ primordial are initiated [120]. KNOX genes in the conifer, *Picea abies*, are expressed in cone buds, needles, roots, stems and embryogenic cultures [72]. *Picea abies* HBK1 (homeobox of KNOX class) transcript is expressed in low levels in the central zone of meristem and is absent from the developing needle primordia [70].

Expression of class II genes is more widespread and found in all organs like leaves, stems, inflorescences, meristems and roots; and their specific functions have not been recognized yet [34, 117, 121, 122]. Class II KNOX proteins have the same domain organization (KNOX1, KNOX2, ELK and HD) as class I KNOX genes (see Figure 1.8) and have more sequence conservation amongst themselves and less with class I KNOX proteins. Class II KNOX genes when overexpressed do not show any altered morphology [117, 123]. KNAT3, KNAT4, KNAT5 and KNAT7 are class II KNOX genes from *Arabidopsis* [100, 122]. KNAT3, KNAT4 and KNAT5 may have a role in root development and function of KNAT7 is not known at the time of this writing [124].

1.6.2 Proteins of KNOX family regulate development by interacting with proteins of BELL family and other families

The KNOX domain of KNOX proteins is also known as the MEINOX domain due to the conservation of KNOX sequence in plants with the MEIS sequence in animals [97]. The KNOX proteins are known to homodimerize through the MEINOX domain [93, 98, 125]. The MEINOX domain of the MEIS proteins has been shown to interact with another family of TALE class called the PBX family [112, 126-128]. In plants, the MEINOX domain of the KNOX proteins has been shown to interact with the BELL family of TALE proteins [93, 100-105, 125]. The interaction of MEINOX with BELL family members is mediated by a conserved domain in the N terminus, called the MEINOX interacting domain (MID), which is composed of SKY and BELL domains [102]. The KNOX-BELL interaction of plant HD proteins parallels the MEIS-PBC interaction in animals. The KNOX-BELL and MEIS-PBC genes have been proposed to share a common ancestor based on their evolutionarily conserved structure and interaction [129]. Studies have shown selective interaction between the KNOX and the BELL family proteins which means that each KNOX protein interacts with a specific subset of BELL proteins and vice versa. KNOX-BELL interactions control floral specification, internode patterning and the maintenance of boundaries between initiating floral primordial and inflorescence meristem. Heterodimerization of KNOX and BELL proteins have been demonstrated in monocots and dicots using various methods such as yeast two-hybrid systems [93, 100-102, 104, 105, 125], in vitro binding [93, 100-103], ligand blot analysis [103] and in living plants cells [105]. In *Arabidopsis*, a BELL protein called PENNWISE (PNY) has been shown to interact with STM and KNAT1 but not

with KNAT2 [103-105]. POUNDFOOLISH (PNF), a BEL1-like homeobox gene which is paralogous to PNY has also been shown to interact with STM and KNAT1 [130]. Another group has shown that BELLRINGER (BLR) binds to STM and KNAT1 but not to KNAT4 [104]. BEL1 has been shown to bind STM, KNAT1, KNAT2 and KNAT5 but to a lesser extent if at all with KNAT3 [100]. VAMANA (VAN) a BEL1-like HD protein has been shown to interact with STM, KNAT1 and KNAT6. Interaction of KNAT1 with PNY and PNF is required for internode patterning and inflorescence development and that of STM with PNY and PNF is required for meristem maintenance [103-105, 130-132]. KNAT1 interaction with BLR, PNY and VAN is required to regulate inflorescence architecture or phyllotaxis in *Arabidopsis* [103-105].

Arabidopsis thaliana ovate family proteins (AtOFPs) have been recently identified and are also known to interact with the KNOX family proteins. These interactions are mediated by the only conserved domain of AtOFP with the HDs of KNOX and BELL proteins [125]. AtOFPs are important regulators of plant development and they control the intracellular localization of TALE family proteins.

KNOX genes are significant to development and their expression is strictly controlled. KNOX protein-protein interaction is key to this control. Variation in protein-protein interaction can have major consequences. Deletion studies in which domains of KNOX regions have been removed, exchanged or mutated reveal a variety of phenotypes. Nagasaki et al. (2001) have shown by creating deletion mutants that KNOX1 has suppressive function and KNOX2 is required for dimerization. If KNOX2 domain or both KNOX1 and KNOX2 domains are deleted from a protein, it cannot form dimer; but if KNOX1 or KNOX2 or both KNOX1 and KNOX2 domains are deleted from a protein, it

can still bind to DNA. Magnani and Hake (2008) have shown that a HD-lacking KNATM-B protein with complete MEINOX domain is able to interact with proteins of the BELL family (PNY, BEL1, SAW1, SAW2) and proteins of the KNOX family (KNAT1, KNAT3, KNAT4). KNATM isoforms A and C with some MEINOX domain sequence deletion do not interact with either. MEINOX domain by itself can form heterodimers but needs HD for homodimerization. [93, 98, 100, 102, 113]. Spliced mRNAs produced by the human Meis2 gene have MEINOX domains but contain a partial HD deletion. The proteins translated from these mRNAs are unable to bind DNA but compete with the HD-containing Meis2 proteins for dimerizing protein partners, and thus function as a dominant-negative regulator [115]. DNA binding HD driven protein-protein interaction has been previously observed in metazoan proteins [109, 133].

1.6.3 Intercellular trafficking of KN1

Vital information for position-dependent specification of cell fate is delivered by intercellular communication. The traditional route of cell-to-cell signaling involves the ligand and receptor interaction [134]. A novel mechanism of intercellular communication that has gained importance in the recent years is intercellular trafficking of regulatory mRNAs and proteins through plasmodesmata [135-140]. These signals can function over a short distance or systemically throughout the plant. Intercellular trafficking of RNAs has been shown to play a critical role in physiological and developmental processes in plants [141-145].

The ability of mRNA/protein movement through plasmodesmata was shown by localization studies, observations of movement after transient expression by microinjection, and grafting experiments and chimeras [47, 135, 136, 141, 146, 147]. The

ability of KN1 to act non-autonomously was first observed in 1990's [148, 149] and concrete evidence for the non-autonomy was obtained from comparisons of protein and RNA localization [47, 48, 140]. Even though zone-specific or cell layer-specific expression patterns have been observed for KNOX mRNA, KNOX proteins have been found to accumulate in regions different from where they have been transcribed [47]. In maize ZmKN1 protein has been found in epidermal cell layers even though their mRNA is not present in those cells, suggesting its trafficking ability between different cells. KNOX proteins have been shown to move between different cell layers through the plasmodesmata [135, 138, 150]. The region containing NLS has been shown by mutational analysis and microinjection experiments to be critical for cell-to-cell trafficking. Using trafficking assay it was shown that KN1 HD is necessary and sufficient for cell-to-cell trafficking. Using the same experiments it was also shown that KNAT2, 3 and 6 do not have trafficking abilities, suggesting that trafficking of knox genes is selective [151]. Grafting experiment was performed in tomato where the wild-type plant was grafted over a knox mutant and due to long-distance movement, the mutant mRNA was found in the wild-type parts of the plant [142]. KN1 protein movement in *Arabidopsis* has been shown by creating KN1-GFP (green fluorescent protein) chimera [138, 150]. In SAM the trafficking of KNOX proteins between inner and outer cells is bi-directional but in leaves they can only move from inner to outer layer i.e. from mesophyll to epidermis but not the other way. These results suggest that intercellular transport plays a critical role in the normal development of KNOX proteins [150].

1.6.4 KNOX genes and Hormones

Plant growth and development are controlled by signaling molecules called hormones. Hormones are synthesized at a particular developmental stage or in particular cell or tissue type and they elicit specific gene expression responses resulting in alteration of cell fate or function. In higher plants, several hormone-associated pathways are intertwined with expression and function of KNOX genes [51].

Auxin – Auxin promotes cell division and cell growth. Antagonistic relationship mainly with respect to organogenesis has been shown between expression and function of knox gene and accumulation of auxin. When auxin transport was inhibited in tomato shoot apex, expression of a knox gene *LeT6* was highly increased [152]. Overexpression of rice *osh1* gene in tobacco causes decrease in levels of auxin [153]. Due to coordinated effects of overexpressed knox genes and reduced auxin levels at the shoot apex, apical dominance is reduced. *stm* mutants show aberrant spatial positioning of leaf primordia due to increased auxin levels [57, 58]. It has been suggested that in the *semaphore1* mutant, which is a knox gene regulator found in maize, ectopic knox gene (*rs1* and *knox4*) expression and reduced polar auxin transport are linked [154].

Gibberellic acid – GA controls cell elongation. A general antagonism has also been shown to exist between knox genes and gibberellic acid (GA) due to its incompatibility with meristematic homeostasis [155]. Transcription of a GA synthesis gene, *ntc12*, which encodes GA-20oxidase, has been shown to be repressed by a KNOX protein, *NTH15*, in tobacco [156, 157]. As a result GA can be synthesized in leaf primordia where there is no KNOX expression, but is suppressed in the SAM where knox genes are expressed. Ectopic expression of rice homeobox gene, *osh1*, causes a decrease in GA levels [153].

Of the four GA synthesis genes in rice (OsGA20ox1-4), OSH15 has been shown to reduce the expression of OsGA20ox2 and 4 [158]. Similar results have also been shown in *Arabidopsis* where expression of GA-20oxidase in the SAM is negatively regulated by STM. *Arabidopsis* leaves show ectopic meristematic activity due to an increase in KNOX activity when GA signaling or biosynthesis is reduced. On the other hand, high GA levels antagonize KNOX function when ectopically expressed in SAM and leaves due to its increased signaling. The abnormalities caused by ectopic knox gene expression in *Arabidopsis* can be partly alleviated by exogenous application of GA [159]. Overexpression of a KNOX protein, *POTH1*, in transgenic lines reduces GA 20-oxidase1 transcript levels in leaves [160].

Cytokinins – Cytokinin promotes cell division and shoot morphogenesis. Cooperative interaction has been shown between cytokinins (trans-zeatin and isopentenyladenine) and knox gene expression. Expression of knox genes increases with the increase in cytokinin levels [95, 153, 161, 162] and plants overexpressing knox genes show increased cytokinin levels [161, 163, 164]. It has been suggested that knox gene function is a critical downstream component of the cytokinin-dependent pathway that promotes shoot development [58]. Overexpression of rice *osh1* gene in tobacco causes an increase in cytokinin expression [153]. Seven IPT synthesis genes have been recognized in *Arabidopsis* AtIPT1 and AtIPT3-8, and knox gene STM has been shown to induce the expression of AtIPT7 and AtIPT1. Eight IPT synthesis genes have been recognized in rice OsIPT1-8 and knox gene OSH1 has been shown to induce the expression of OSIPT2 and 3 [158].

Ethylene – Ethylene controls fruit ripening. Antagonistic relationship has been shown between knox gene function and ethylene. When a precursor of ethylene, ACC is applied to plants, the phenotypes due to KNAT2 overexpression can be suppressed [165].

Abscisic acid – ABA inhibits cell division. ABA levels are also shown to be regulated by knox genes. Overexpression of the rice *osh1* gene in tobacco causes ABA levels to increase in tobacco plants [153].

1.7 KNOX genes and Evolution

A key player in genome evolution is gene duplication and it can take place by whole genome or chromosomal duplication, unequal crossing over or retroposition.

Duplicated genes form gene families and are known as paralogous genes [166]. The fate of a duplicate gene is either to conserve its function (gene conversion), acquire a novel function (neofunctionalization) or adopt part of the function from the parental gene (subfunctionalization). Examples of gene conversion are ubiquitin and histone genes which are required in large amounts [167, 168]. An example of neofunctionalization is eosinophil-derived neurotoxin which after duplication acquires new function and is called eosinophil cationic protein [169]. An example of subfunctionalization is engrailed gene which is a single gene in mouse, but the same function is performed by two different genes engrailed-1 and engrailed-1b which are generated by duplication in zebrafish [170, 171]. If duplicated genes are not under selection pressure, after few million years they become pseudogenes [172].

Occasionally mRNAs are reverse transcribed and integrated back in the genome (retroposition) with the help of trans-acting polymerase and integrase respectively. The resulting structures are known as processed pseudogenes. They lack promoter and

introns, and may have a poly-A tail and possibly reading frame disruptions due to frame shifts or premature stop codons [173]. Once formed, they integrate themselves into a location different from the original gene [166]. Some examples of pseudogenes described in plants are the actin pseudogene in potato [174], hmg-coA reductase pseudogene in cotton [175], the ARI pseudogenes in *Arabidopsis* [176]; the cdc2 pseudogene in Norway spruce [177]. Three hundred seventy six processed pseudogenes have been reported in *Arabidopsis* [178].

It has been suggested that class I and II knox gene families have been formed by gene duplication that took place before the divergence of Bryophytes (mosses) and Spermatophytes (seed plants) which occurred at least 400 MYA [179]. Gymnosperms and angiosperms are thought to have diverged from a common ancestor about 300 MYA [180, 181]. Within the gymnosperms, the conifers are thought to have diverged about 140 MYA [182]. Class I knox genes in conifers (KN1, KN2, KN3, KN4) are thought to have diverged about 160 MYA before the split between pine and spruce [96]. In angiosperms, KNOX-I genes have diversified into four groups (A1, A2, A3 and A4). Gymnosperms have only one of these groups which is sister to A4 and all the class I knox genes known in conifers so far belong to A4 and have diverged into four subgroups (KN1, KN2, KN3 and KN4). These conifer sequences form a monophyletic group and are not shared by angiosperms. Thus class I knox genes in conifers evolved differently from those in angiosperms probably by gaining new paralogs and losing the common genes after their split [96]. The Loblolly pine genome is very large and complex (approximately 2×10^{10} bp; roughly 100 fold larger than the *Arabidopsis* genome). It is composed of highly

repetitive sequences; and multiple copies of genes which could be either functional gene or pseudogene are generated by gene amplification [183].

1.8 Altered activity of a single gene has profound effects on development

Very different morphologies, life cycles and characteristics of plants can be explained by the difference in expression of regulatory genes which are mostly caused by mutations in promoters and enhancers [184]. An example of this is the development of maize as domestic crop from its progenitor teosinte by change in expression of a single gene teosinte branched1 (tb1) which increased apical dominance [185]. Another example is the development of a large tomato fruit from its small progenitor due to change in regulation of a single gene fruitweight2.2 (fw2.2) [186]. An example of change in phenotype due to change in single amino acid is of a myb transcription factor gene shattering4 (sh4) in rice; which causes the wild rice to fall on ground and the cultivated rice to stay on the plant [187]. See Figure 1.10 for the altered phenotypes of these examples.

A suite of genes is required for correct embryo development, but alteration in a single gene can have profound effects on development. Alteration in the expression of knox genes, through mutation or by increasing or decreasing mRNA levels in transgenic plants, produces profound changes in morphology and growth. For example, mutation or overexpression of knox genes in transgenic plants causes great difference in size, shape and development [50, 51]; therefore is an important gene and could provide clues to embryo development in gymnosperms and angiosperms. In the following chapters I describe and discuss knox gene and its novel spliced forms in gymnosperm and angiosperm.

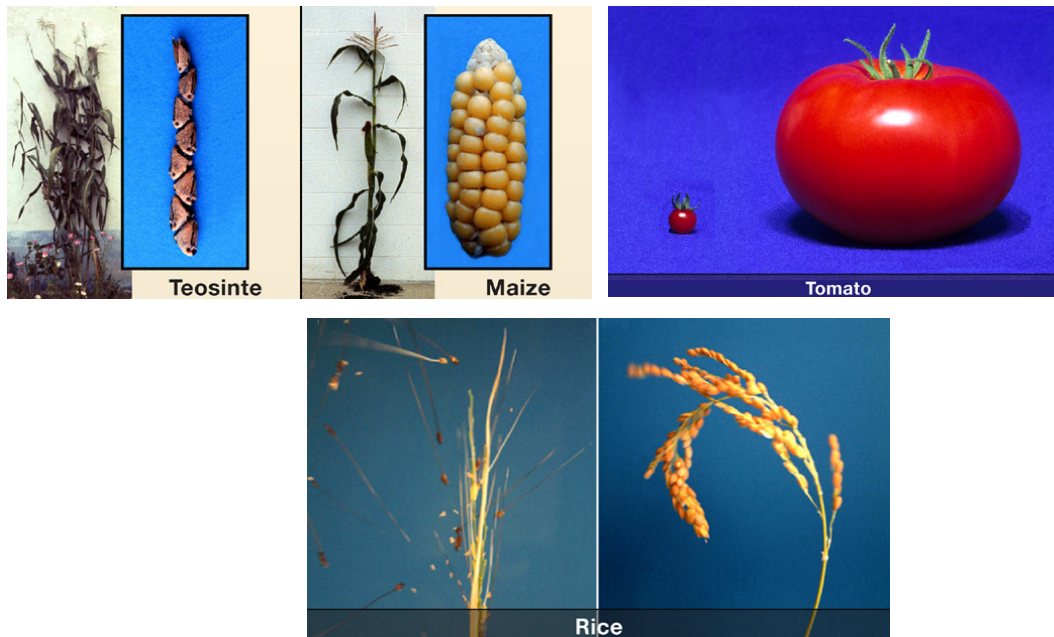


Figure 1.10: Altered activity of a single gene has profound effects on development [188].

Change in expression of *ts1* causes domesticated maize to form from wild-type teosinte.

Change in regulation of *fw2.2* causes tomato size to increase 30%.

Change in a single amino acid of *sh4* converts shattering rice to become non-shattering.

CHAPTER 2

**KNOX PROTEINS LACKING A HOMEODOMAIN ARE
PRODUCED VIA ALTERNATIVE SPLICING OF KNAT-1 LIKE
GENES IN GYMNOSPERMS AND ANGIOSPERMS**

2.1 Introduction

Embryogenesis in Gymnosperms and Angiosperms while broadly similar differs in important respects. In pines the zygote undergoes several rounds of nuclear duplication before forming walls to create mononuclear cells; multiple embryos develop initially before one dominates and ultimately a single elongate embryo with 6-8 cotyledons develops. All of the above unfolds within a haploid gametophyte, the endosperm being absent [28, 30, 189]. In *Arabidopsis*, the zygote divides asymmetrically to establish a smaller apical cell which produces the embryo, and a larger basal cell which will form the suspensor; the embryo assumes a variety of shapes on the way to forming a dicotyledonous structure, folded over within a triploid endosperm [26, 190]. Despite these differences, the genes expressed in developing pine embryos show strong sequence similarity to those found in angiosperm genomes [31]. The distinguishing features of conifer embryogenesis may be a consequence of the level, timing or localization of expression of certain regulatory genes perhaps supplemented by the activity of a few unique or highly differentiated proteins.

The KNOX family of regulatory genes exerts considerable influence upon plant development [50, 51]. Alteration in the expression of KNOX genes, through mutation or by increasing or decreasing mRNA levels in transgenic plants, produces profound

changes in morphology and growth. The ability of these proteins to affect a broad range of cellular activities is explained in great measure by the presence within the KNOX proteins of distinct domains that can interact with a variety of other regulatory proteins. The HD region of KNOX proteins is a sequence-specific DNA-binding domain that can directly regulate expression of target genes [77, 78]. The KNOX1 and KNOX2 regions constitute the MEIKNOX domain [97] which has been shown to homodimerize with KNOX family proteins [98, 114] or heterodimerize with members of the BELL family of TALE HD proteins [93, 100-105]. However the HD itself is capable of mediating protein-protein dimerization [98, 114] possibly via the ELK domain [34, 106]. Individual KNOX proteins appear to interact with BELL proteins in a discriminating manner, with certain interactions being strong while others are much weaker [100, 103, 104]. KNOX genes have been identified in conifers, mRNAs have been cloned and the predicted proteins have been shown to possess features common to the KNOX family [70, 71, 96, 191]. Here we report a novel conifer KNOX mRNA which lacks HD sequences and is produced by cleavage-polyadenylation within the large third intron of the PtKN1 gene. This transcript and its encoded protein are expressed during embryogenesis. We show comparable mRNAs and proteins are produced by the *Arabidopsis* KNAT1 gene, and by the maize RS1 gene.

2.2 Results

2.2.1 PtKN1 – the KNAT1-like mRNA of Loblolly Pine (*Pinus taeda* L.) Contains Highly Conserved Domains

By employing primers based on the sequence of the Norway spruce (*Picea abies*) KNAT1-like mRNA (HBK3, AF483278) in RT-PCR, I cloned fragments of mRNAs

expressed at different stages of embryo development in loblolly pine (*Pinus taeda* L.). Using these sequences, 5`RACE was used to obtain the 5` end of the mRNAs. Finally 3` RACE was used to generate single-molecule cDNA clones containing the 5`UTR sequences, the complete protein-coding sequence and regions of the 3` UTR (Supplementary Data Figure A.1). The open reading frame of these clones was 963 nucleotides long and, due to differences in the length of the 3`UTR, the cDNA clones varied in length from 1400 to 1900 nucleotides. Taking the longest clone as reference, the sequences of twenty-nine independently isolated PtKN1 cDNAs were compared and shown to be virtually identical, with a minimum of zero to a maximum of 8 nucleotide differences over their length. The predicted protein from the pine KNAT1-like mRNA (PtKN1) shows strong conservation of sequence within the KNOX1, KNOX2, and HD regions, when compared to *Arabidopsis* (Figure 2.1).

```

Pt KN1      MNVERLNTAAACASSSLYGVNMAEYGDAGVSSMGLI TCHPEECSENI MSTRIPSSFSF
KNAT1      -----MEEY-----CHDN----STTPCRVSYLYSPI
consensus  mnmerlntaaqassslygvnm- EYgdagvssrmglt tChe- eqse-----Ri---fS--

Pt KN1      HGHADCLLSAAMFQGSQGDHKLNPQPGMNCQLVSECSI MSDSSMPLVKT KACSGLRNCFE
KNAT1      SSSNKNDNISDTNNNNNNSSNYGPGYNTNNN- NHHQHMLFPHVSSLLPCTTENCER
consensus  -----s-----q--q-----N--PG-Nq-----eq-----P-v-t-----N-F-

Pt KN1      FHRECPGNCYTDGSSNI PL SPI V TSLASCARGEARMI PSLDANSAHFVNDNEHAHKSKI
KNAT1      SDHDCPNNNNNPSVKSEASSRI NHYSMLMR---A HNTCEANNNNNNDNVSIVEANKAKI
consensus  -----eCP-N-----S--v-----Rgea-r--s-dAN-----e--AiK-KI

Pt KN1      LAHPCYPSLLGAYLDCCKI GAPPEAVARLDALTRHCDPQRR--TVSI G-MPELDQFM
KNAT1      LAHPHYSILLCAYLDCCKI GAPPEVVDRI TAARCFEARGCRSTPSVSASSRDPELDQFM
consensus  LAHP-Y-sLL-AYLDCCKI GAPPe-V-RI-A---e---G-RstptVS--s-DPELDQFM

Pt KN1      EAYCEILTKYHEELAKPFKEAMLFKKI ETQFNSLGKGTI RISSPADDEKTEGGGSSEE
KNAT1      EAYCDMLVKYREELTRPI GEAMEFIRRI ESQLSMLCGSPI HILNNDGKSDNMGSSDEE
consensus  EAYCeIL-KY-EEL-kP--EAM-FI kki Et Q---L-----I-I-----D-----G-gS-EE

Pt KN1      VEDGSGGETDFCEVDHHAIVEDRELKNHLLRKYSGYLSLTKGEFMKKKKKGKLPKQARQKL
KNAT1      QENNSGGETELPEIDPRAE-DRELKNHLLRKYSGYLSLTKCEL SKKKKKKGKLPKQARQKL
consensus  -E--SGGETd--EvD--A-eDRELKNHLLrKYSGYLSLTKGE--KKKKKGKLPKdARQKL

Pt KN1      LDWMSLHDKWPYPSETEKIALAECTGLDCKGI NNWFI NCRKRHWKPSMDMFMVMN-SHS
KNAT1      LITWELHYKWPYPSESEKVALAESTGLDCKGI NNWFI NCRKRHWKPSMDMCFMMDGLCH
consensus  L-WW-LH-KWPYPSEtEKI ALAE-TGLDCKGI NNWFI NCRKRHWKPSMDV-FVMV-g---

Pt KN1      PHSAALYVERHMMTEG-YL---
KNAT1      PHHAALYMDGHYMGDGPYRLGP
consensus  PH-AALYve-H-M-eGpY-lgp

```

Figure 2.1: Alignment of the amino acid sequences deduced from the sequences of PtKN1 and Arabidopsis KNAT1 (NM_116884) mRNA.

There is overall amino acid identity of 62% (86% in HD); and similarity of 77% (92% in HD)) between the sequences. Conserved domains are indicated as follows: green - KNOX1; yellow - KNOX2; blue - ELK; red – HD.

2.2.2 A KNAT1-like mRNA (PtKN1(hd-)) Lacking HD Sequences is Expressed

During Pine Embryo Maturation

Unexpectedly, I isolated several clones of a PtKN1 mRNA which encode a putative protein containing the MEINOX domain but lacking the HD (Figure 2.2, Supplementary Data Figure A.2). I will, henceforth, refer to mRNAs containing HD sequences as (HD+) and those lacking such sequences as (hd-).

The nucleotide sequence of PtKN1(hd-) mRNA is identical to the PtKN1(HD+) mRNA sequence from the 5'UTR through the KNOX1 and KNOX2 domains, thereafter a novel sequence is present containing a stop codon which terminates the PtKN1(hd-) protein 39 amino acids after the point at which the (HD+) and (hd-) sequences diverge (Figure 2.2). The PtKN1(hd-) mRNA contains a lengthy 3'UTR which bears no similarity to any previously identified sequence. The domain structure of the two mRNAs is shown diagrammatically in Figure 2.3 and the nucleotide alignment is given in Supplementary Data Figure A.3). This is the first identification of a plant KNAT-1-like mRNA which lacks HD sequences.

```

Pt KN1 [ HD+] MNMERL NTA AACASS SLYGVNMAEYGDAGVSSM/GLI TCHEPECS ENI MSTRI PSSFSSF
Pt KN1 [ hd-] MNMERL NTA AACASS SLYGVNMAEYGDAGVSSM/GLI TCHEPECS ENI MSTRI PSSFSSF
consensus MNMERL NTA AACASS SLYGVNMAEYGDAGVSSM/GLI TCHEPECS ENI MSTRI PSSFSSF

Pt KN1 [ HD+] HGHADCLL SAAMFCGSCGDHKLNP CPGMNGCLVSECSI MSDSSMPLVKT KACSGLRNCFE
Pt KN1 [ hd-] HGHADCLL SAAMFCGSCGDHKLNP CPGMNGCLVSECSI MSDSSMPLVKT KACSGLRNCFE
consensus HGHADCLL SAAMFCGSCGDHKLNP CPGMNGCLVSECSI MSDSSMPLVKT KACSGLRNCFE

Pt KN1 [ HD+] FHRECPGNCYTDCSSNI PLSP I V TSLASCARGEARM I PSLDANSAHFNV DNEEHAI KSKI
Pt KN1 [ hd-] FHRECPGNCYTDCSSNI PLSP I V TSLASCARGEARM I PSLDANSAHFNV DNEEHAI KSKI
consensus FHRECPGNCYTDCSSNI PLSP I V TSLASCARGEARM I PSLDANSAHFNV DNEEHAI KSKI

Pt KN1 [ HD+] LAHPCYPSLL GAYI DCCKI GAPPEAVARLDAL TREHCDP GRRTVSI GMDPELDGFMEAYC
Pt KN1 [ hd-] LAHPCYPSLL GAYI DCCKI GAPPEAVARLDAL TREHCDP GRRTVSI GMDPELDGFMEAYC
consensus LAHPCYPSLL GAYI DCCKI GAPPEAVARLDAL TREHCDP GRRTVSI GMDPELDGFMEAYC

Pt KN1 [ HD+] E I L T K Y H E E L A K P F K E A M L F L K K I E T Q F N S L G K G T I R I S S P A D D D E K T E G G G S S E E V E D G
Pt KN1 [ hd-] E I L T K Y H E E L A K P F K E A M L F L K K I E T Q F N S L G K G T I R I S S P A N T R L L H K H F L L L F S P I A P
consensus E I L T K Y H E E L A K P F K E A M L F L K K I E T Q F N S L G K G T I R I S S P a -----

Pt KN1 [ HD+] S G G E I D F C E V D H H A V E D R E L K N H L L R K Y S G Y L S S L K C E F M K K K K K G K L P K D A R G K L L D W W
Pt KN1 [ hd-] I N S A S Y C L H Y V N S R Y S G D K T -----
consensus -----t-----knhl l r k y s g y l s s l k q e f m k k k k k g k l p k d a r q k l l d w w

Pt KN1 [ HD+] S L H D K W P Y P S E T E K I A L A E C T G L D C K Q I N N W F I N C R K R H W K P S E D M H F M V M N S H S P H S A A
Pt KN1 [ hd-] -----
consensus s l h d k w p y p s e t e k i a l a e c t g l d q k q i n n w f i n q r k r h w k p s e d m h f m v m n s h s p h s a a

Pt KN1 [ HD+] L Y V E R H M M T E G Y L
Pt KN1 [ hd-] -----
consensus l y v e r h m m t e g y l

```

Figure 2.2: Alignment of the amino acid sequences deduced from the sequences of PtKN1(HD+) and PtKN1(hd-) mRNA.

There is 100% identity between the sequences in the overlapping region. Conserved domains are indicated as follows: green - KNOX1; yellow - KNOX2; blue - ELK; red - HD.

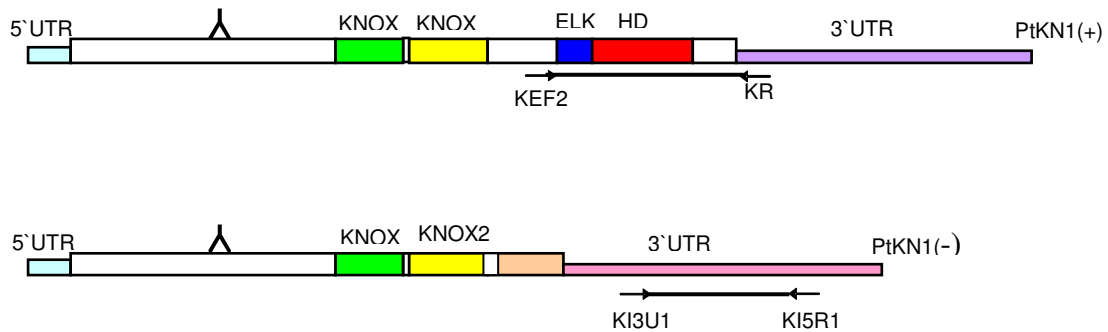


Figure 2.3: Domain structure of PtKN1(HD+) and PtKN1(hd-) mRNA.

Thick boxes represent the protein-coding regions and the thin boxes represent the untranslated regions.

Primers used for the RT-PCR analysis – Location of gene specific primers KEF2+KR and KI3U1+KI5R1, used for semi-quantitative PCR are shown above. These primers generate 472 bp and 391 bp products for PtKN1(HD+) and PtKN1(hd-) forms respectively.

Position of antibody binding - A polyclonal antibody was raised against a peptide sequence which is present in both the PtKN1 forms. The location of the epitope, the site of the antibody binding is indicated by a 'fork' above the sequence.

Conserved domains and untranslated regions are indicated as follows: turquoise – 5' UTR; green - KNOX1; yellow - KNOX2; blue - ELK; red - HD; tan - (hd-) specific ORF sequences; pink - (hd-) 3' UTR; purple - (HD+) 3' UTR.

2.2.3 The PtKN1(hd-) mRNA is Expressed Abundantly During Pine Embryo

Maturation

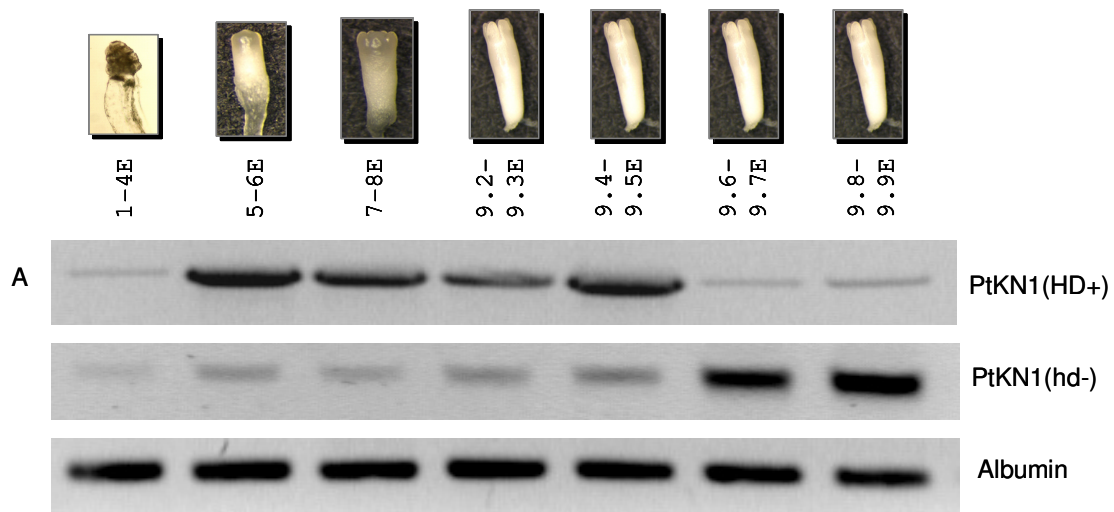
Expression of a KNAT1-like protein which lacks a HD could have significant influence upon development, either by suppressing KNOX activity through forming non-functional dimers with HD-containing proteins or through novel interactions resulting in altered gene regulation. To determine the expression pattern of these transcripts during embryo development, semi-quantitative PCR was performed using primers specific for the two forms of mRNA; primer locations are shown in Figure 2.3. Figure 2.4A shows that the PtKN1(HD+) transcript is expressed fairly steadily throughout pine embryo development. By contrast, the level of PtKN1(hd-) mRNA rises gradually, peaking during the late maturation phase of development, after the embryo has reached its final shape and size. At its zenith, the abundance of PtKN1(hd-) mRNA greatly exceeds that of PtKN1(HD+) mRNA and is virtually the exclusive form of KNAT1-like mRNA present in the embryos immediately prior to germination. Albumin has been shown to be successfully used as control in semi-quantitative PCR in loblolly pine [192] and therefore was used as control in this research.

2.2.4 PtKN1(hd-) Protein is More Abundant than the PtKN1(HD+) Protein in Late Maturing Pine Embryos

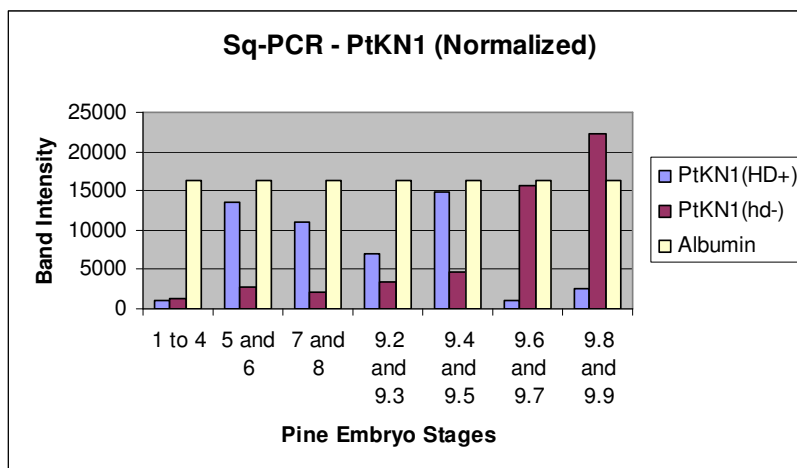
To establish that these mRNAs are translated, antibodies were generated against a loblolly pine PtKN1 sequence common to both forms of protein (Figure 2.4B, Supplementary Data Figure A.4). The expression patterns observed were broadly similar to those observed for mRNA. The amount of PtKN1(hd-) protein rose over development, peaking during late embryo maturation. The quantities of PtKN1(HD+) protein varied

little over development and did not fall away in late development, as was observed for PtKN1(HD+) mRNA. One notable feature is that, in late maturing embryos, the PtKN1(hd-) protein was considerably more abundant than the HD-containing version, consistent with the quantitative differences in mRNA observed earlier (Figure 2.4A).

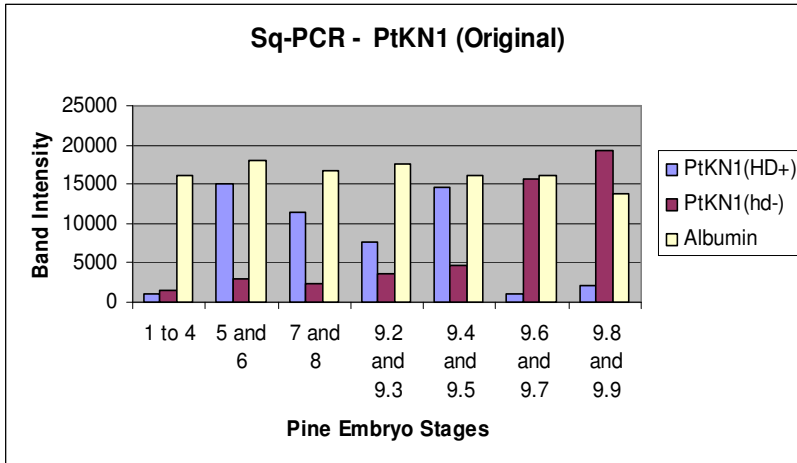
SUMMARY FIGURE SHOWING RELATIVE EXPRESSION OF HD+ and hd-



A : Enzyme - Advantage2 Taq Polymerase
 No. of cycles - 30 for hd- and 35 for HD+ and albumin
 Volume loaded on gel - 8ul each

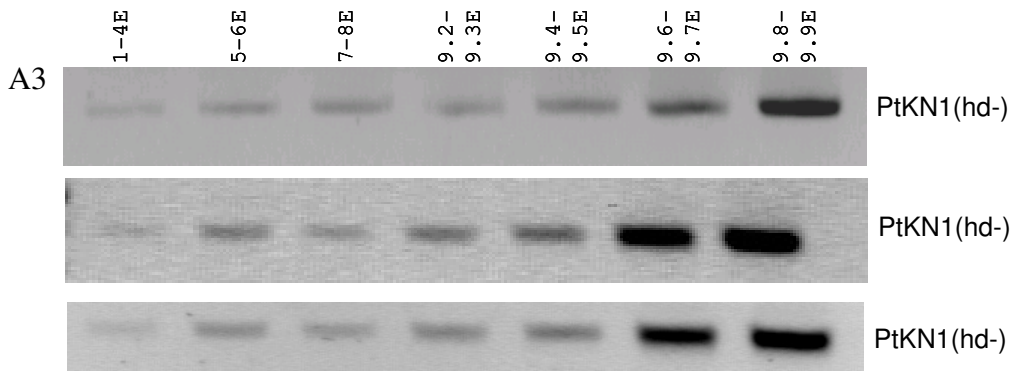


A1 - Normalized graph of Sq-PCR for PtKN1(HD+), PtKN1(hd-) and Albumin



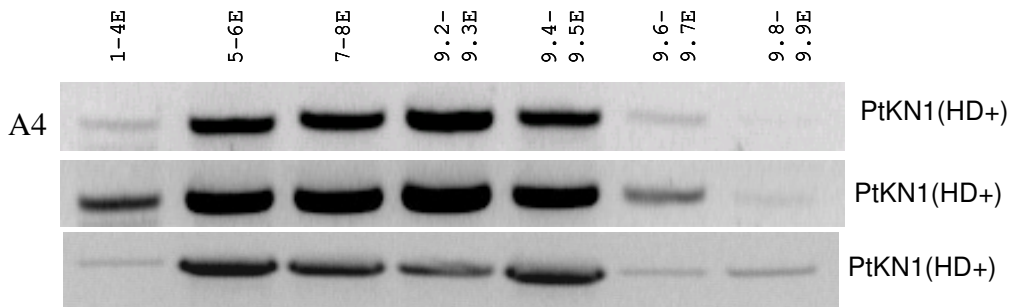
A2 - Original graph of Sq-PCR for PtKN1(HD+), PtKN1(hd-) and Albumin

THREE INDEPENDENT spPCR EXPERIMENTS for (hd-)



A3 - Enzyme – Advantage 2 Taq polymerase
 No. of cycles - 30
 Volume loaded on gel - 8ul each

THREE INDEPENDENT spPCR EXPERIMENTS for (HD+)

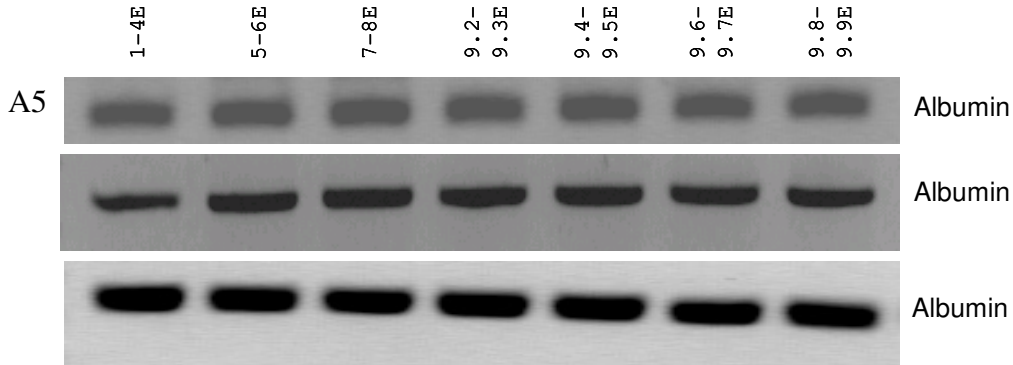


A4 - Enzyme – Advantage 2 Taq polymerase

No. of cycles – 35

Volume loaded on gel - 8ul (top and bottom panel) or 12 ul (middle panel)

ALBUMIN CONTROLS:

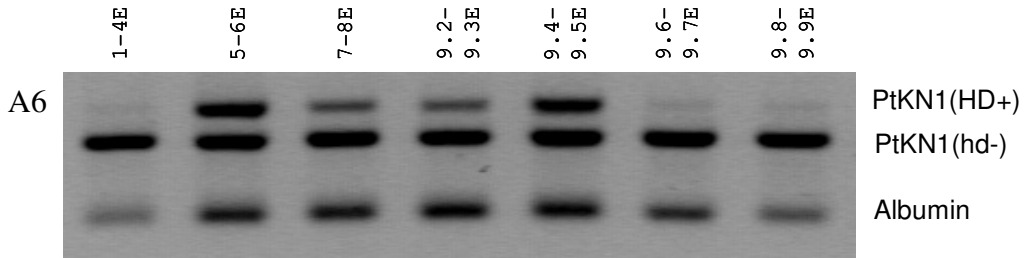


A5 - Enzyme – Advantage 2 Taq polymerase

No. of cycles – 35

Volume loaded on gel - 8ul each

sqPCR USING GREEN TAQ ENZYME



A6 - Enzyme - Green Master Mix

No. of cycles - 35

Volume loaded on gel - 8ul each

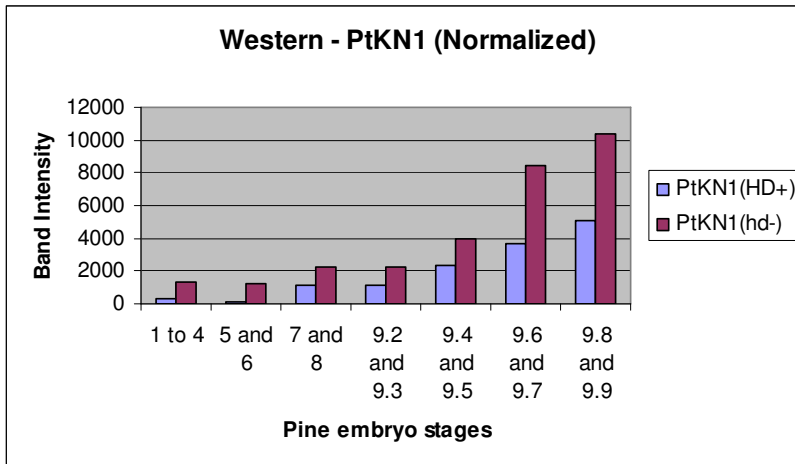
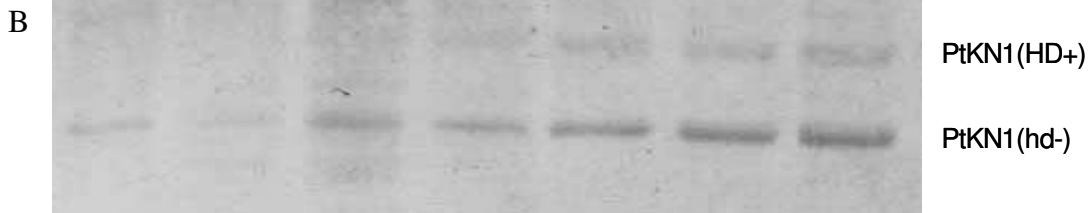
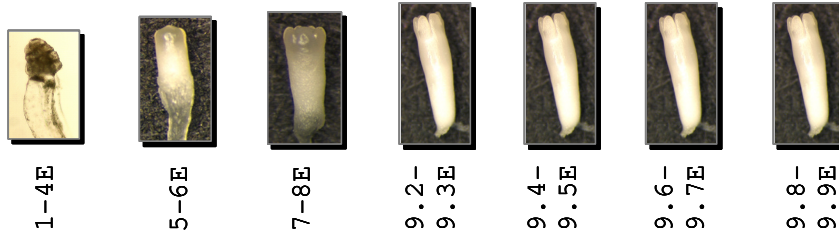


A7 - Enzyme - Green Master Mix

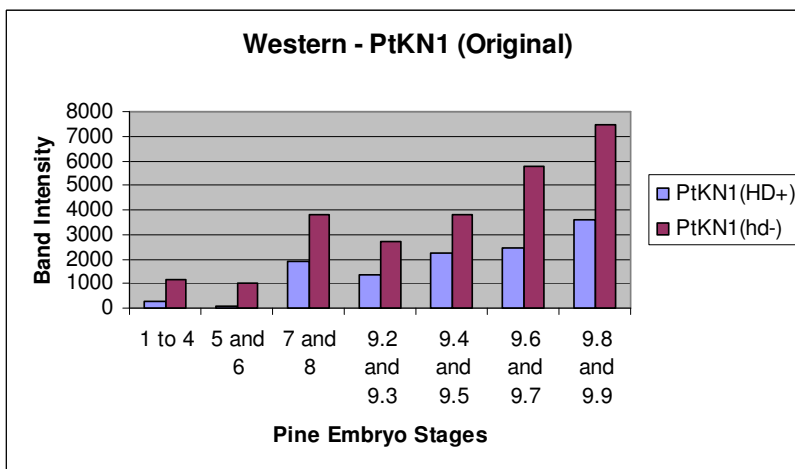
No. of cycles - 35

Volume loaded on gel - 5ul each

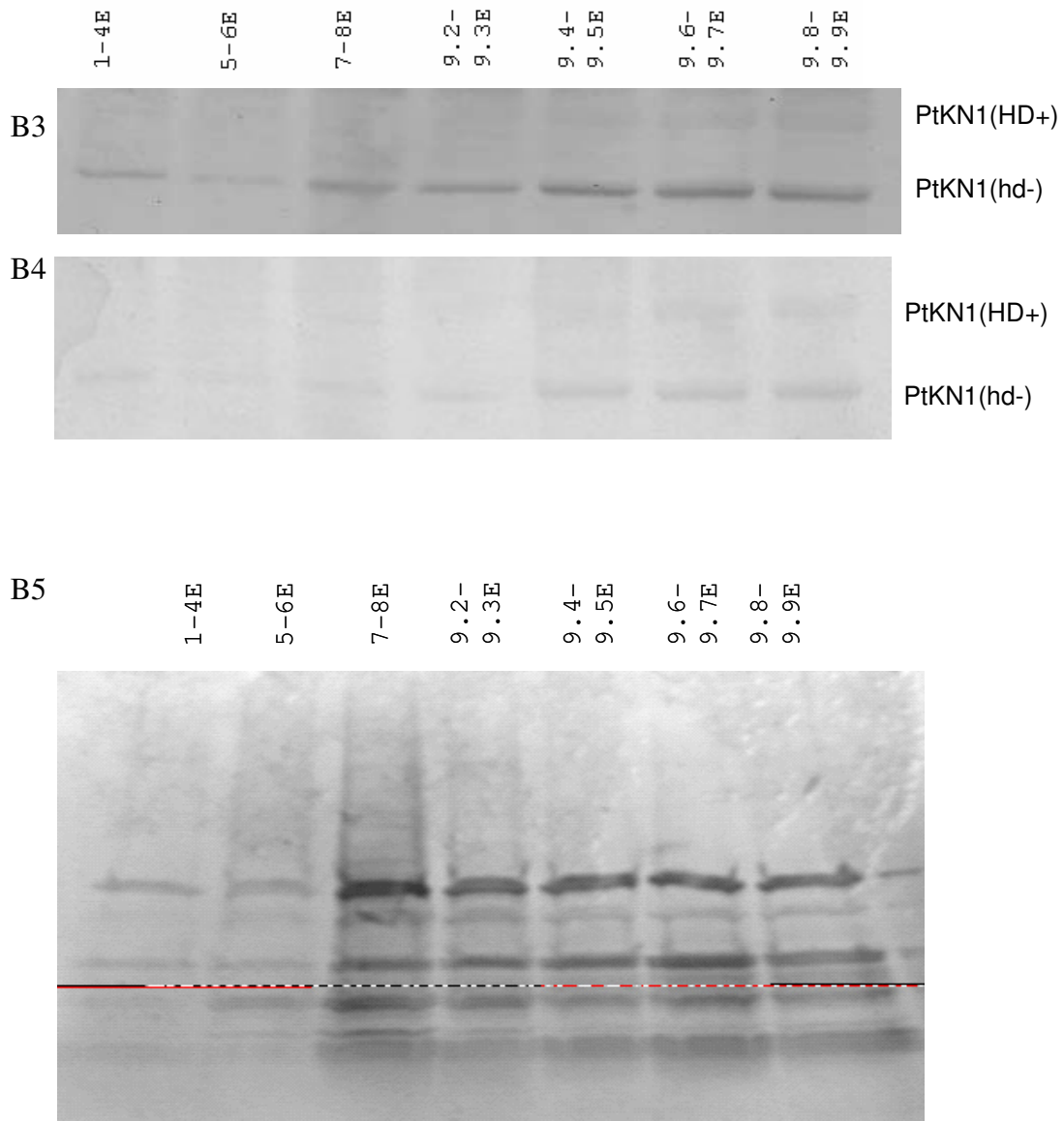
WESTERN BLOTS:



B1 – Normalized graph of Western blotting for PtKN1(HD+) and PtKN1(hd-)



B2 – Normalized graph of Western blotting for PtKN1(HD+) and PtKN1(hd-)



Control protein loading for Western blotting.
12ug of protein loaded for each lane.

Figure 2.4: PtKN1(HD+) and PtKN1(hd-) mRNAs and proteins are developmentally regulated.

(A) Semi-quantitative RT-PCR for pine embryos over development. RNA was extracted from pine embryos at different stages of development, as shown above the lanes and semi-quantitative PCR was performed. 1- stages 1-4; 2- stages 5-6; 3- stages 7-8; 4- stages 9.2-9.3; 4- stages 9.4-9.5; 4- stages 9.6-9.7; 4- stages 9.8-9.9. Albumin was used as control. For staging system see Ciavatta et al. (2001).

(A1-A2) Normalized and original graphs for sq-pcr for PtKN1(HD+), PtKN1(hd-) and albumin.

(A3)-(A7) Replicates of Figure 2.4A

(B) Western analysis of PtKN1 proteins over embryo development.

Protein was extracted from the same pine embryo used for semi-quantitative PCR. Western blotting was performed using antibody raised in the region shown in Figure 2.3. Two proteins were detected one of approximately 48kD and the second of 35kD which correspond to the size obtained from putative protein sequence of PtKN1(HD+) and PtKN1(hd-) respectively based on their cDNA sequences.

(B1-B2) Normalized and original graphs for western blotting for PtKN1(HD+), PtKN1(hd-).

(B3)-(B4) Replicates of Figure 2.4B

(B5) Stained protein gel - loading control for Figure 2.4B, Figure 2.4B3, Figure 2.4B4 Equal volume of protein was loaded for each lane. The gel was stained with simply blue safe stain (Invitrogen, Carlsbad, CA). The loading of the protein does not look even. Normalized graph is true representation of the expression rather than the original where the expression of lane 3 (stage 7-8) is higher because of more protein loading compared to the other stages.

See methods and materials section for the procedure used to make graphs.

2.2.5 Genomic Cloning of PtKN1

The nucleotide identity of the pine PtKN1(hd-) and PtKN1(HD+) mRNAs in their regions of overlap would be expected if both mRNAs were derived by alternative splicing of a PtKN1 RNA from the same gene. In *Arabidopsis*, KNAT1 is a single gene and no splice variant lacking a HD has been reported [60]. In conifers, no complete genomic clone of KN1 has been isolated. This may be due to cloning difficulties caused by the large size of the third intron in many KN1 genes; over 5kb in maize (*Zea mays*), 5.5kb in rice (*Oryza sativa*) and over 4 kb for KNAT2 in *Arabidopsis* [60, 90, 193]. Guillet et al. (2004), cloned genomic regions which were 5' and 3' to intron 3 of KN1 for several conifers but did not clone intron 3.

Using gene specific primers based on the mRNA sequences I generated single molecule clones of the 5' end and 3' end of the PtKN1 gene. I made numerous attempts to generate a single molecule clone which spanned the third intron of PtKN1 however I was unsuccessful; consequently I employed 'Genome Walker' technology (Clontech, MountainView, CA) to obtain overlapping clones of PtKN1 intron 3. A diagram of the

structure of the PtKN1 gene is shown in Figure 2.5. Sequence of the primers used for isolation of genomic sequence is shown in supplementary data table A.1 and location of those primers and a contig map of the PtKN1 gene sequence are shown in supplementary data Figure A.5.

The sequences of PtKN1 exons 1, 2 and 3 are identical to the sequences found in the 5' end of both PtKN1(HD+) and PtKN1(hd-) mRNAs. The boundaries of exon 3 are traditionally defined by alignment of the genomic sequence with the principal mRNA, in this case PtKN1(HD+). However, in pine the genomic DNA, the region which is common to both mRNAs is immediately followed, by the unique sequence found at the 3' end of the PtKN1(hd-) mRNA i.e. the 3' end of the PtKN1(hd-) mRNA is encoded by the proximal end of intron 3. Eleven PtKN1(hd-) genomic clones covering exon 1 through the proximal end of intron 3 showed 99.6% identity (1 to 9 nucleotide differences over 2090 bp). The exonic regions of these clones show 100% nucleotide identity with (HD+) and 100% nucleotide identity with (hd-) cDNAs. These data argue that the PtKN1(hd-) mRNA is derived from transcription and processing of this PtKN1 gene.

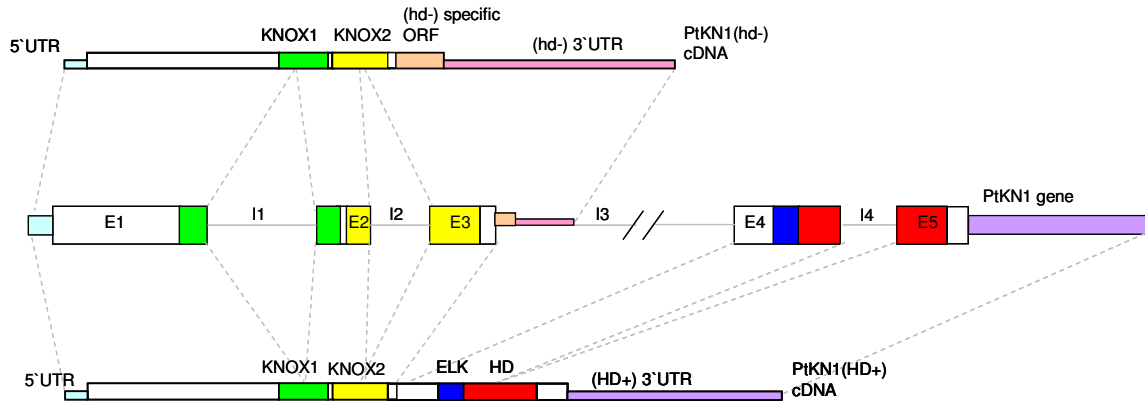


Figure 2.5: A single PtKN1 gene produces two mRNAs – PtKN1(HD+) and PtKN1(hd-).

Diagram of PtKN1 Gene Structure. exons (E) and exons (I) are represented by Boxes and Lines respectively. exons 1, 2 and 3 which encode the region from 5' UTR to KNOX2 domain are common to PtKN1(HD+) and PtKN1(hd-) mRNAs. The 3' UTR of PtKN1(hd-) is encoded by the proximal region of exon3. The PtKN1(hd-) mRNA could be produced by splicing to remove the first two introns followed by cleavage-polyadenylation or termination within intron 3. While spliced removal of exon3 and 4, would exclude the sequences encoding the 3' UTR of PtKN1(hd-) and produce the HD containing, PtKN1(HD+) mRNA.

The 4th and 5th exons encode the HD. A GU-AG splice which removes intron 3, creates the HD+ form of the mRNA, and in so doing excludes the sequence found at the 3' end of the hd- mRNA. The mechanism determining which type of mRNA is produced may be complex. The PtKN1(hd-) mRNA could be produced by splicing of the first two introns followed by cleavage-polyadenylation of the primary transcript emerging from RNA polymerase as it traverses intron 3. The regulatory dynamic at play would be a competition between cleavage-polyadenylation within intron 3, to produce PtKN1(hd-) mRNA, and removal of intron 3 by splicing, to produce PtKN1(HD+) mRNA.

Conifer genomes are large (approximately 2×10^{10} bp); roughly 100-fold larger than the *Arabidopsis* genome with multicopy gene families and processed pseudogenes being common [183, 194]. The sequence identity of the PtKN1(HD+) and PtKN1(hd-)

mRNAs suggests that they originate from a single gene however, in such a large genome, additional forms of the PtKN1 gene, such as a version of the KN1 gene which lacks a HD, may be present. To assay the heterogeneity of the PtKN1 gene family I generated and cloned genomic fragments spanning different regions of the gene using PCR primers in exonic regions. Single genomic fragments of expected size were observed and cloned for analysis. The genomic inserts of individual plasmid clones were amplified by PCR and digested using several restriction enzymes which cut within intronic and exonic regions. Digestion patterns for 14 independent genomic clones of the region from the 5'UTR through exon 2; and the region at 3' end covering exon 4, intron 4 and exon 5 show identical digestion patterns (except clones 8 and 14 in Figure 2.6B and Figure 2.6C, clone 10 in Figure 2.6E and Figure 2.6F; which are consistent with same gene structure in opposite orientation). These data are consistent with a single form of the PtKN1 gene within loblolly pine genome.

A further search for sequence heterogeneity was performed by cloning genomic DNA which lies 5' to the coding region. Using a primer within the ORF which is common to both types of transcript, I performed 'genomic walks'; cloning region 5' of the PtKN1 gene. Sixteen independent genomic clones of 1.0kb in length were sequenced. These clones represent approximately 945 bp of the sequence upstream of the putative TSS. All sixteen clones were nearly identical in sequence. Taking one clone as a reference I observed between 0 and 8 random nucleotide differences over 1.0kb (data not shown). Subsequent isolation of overlapping clones and assembly of contiguous sequence has provided 1,730 bp of 'promoter' sequence. The sequences of eight independent genomic clones representing sequence 1236 – 1669 bp upstream of the putative TSS were

identical. The lack of sequence heterogeneity within 1,730 bp of PtKN1 promoter sequence is consistent with a single form of PtKN1 gene.

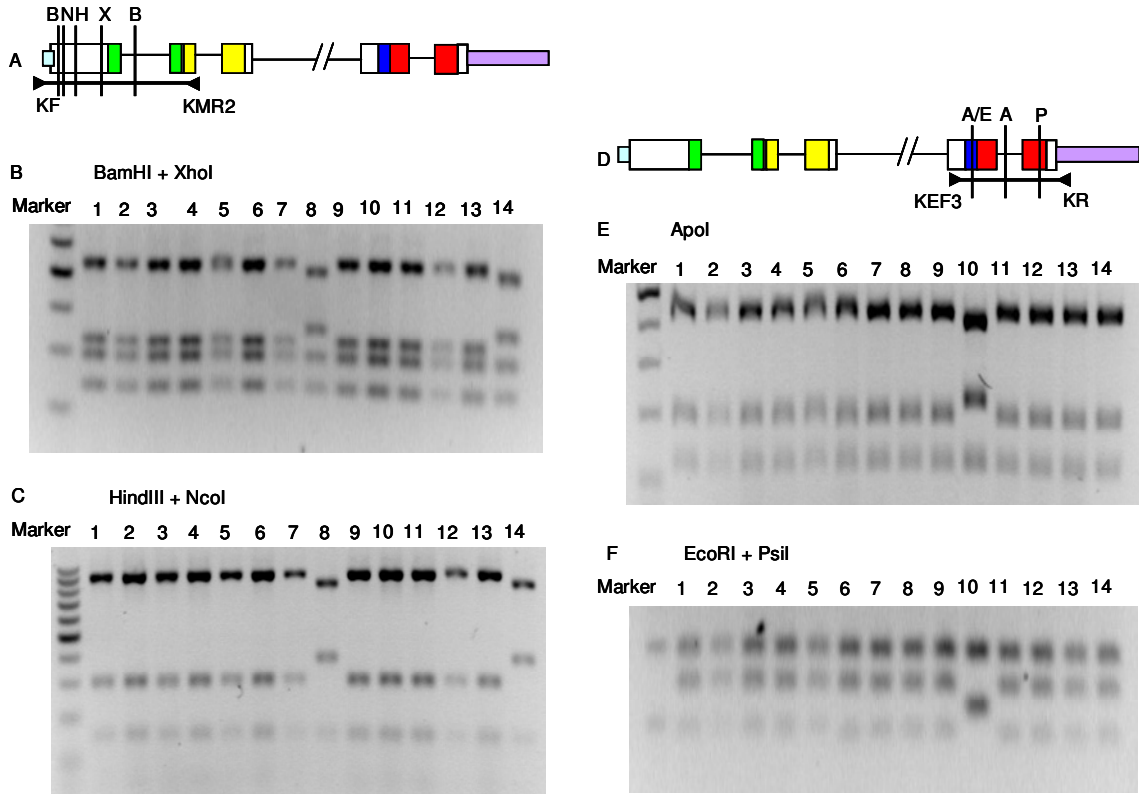


Figure 2.6: Restriction digestion of different regions of genomic DNA to determine PtKN1 structure.

Region covering exon1, intron1 and exon2 and region covering exon4, intron4 and exon5 were amplified with primers as shown in A and D respectively. 14 independent clones of each region were digested with restriction enzymes that cut at locations marked by vertical bars as shown in A and D. B and C are digestion of clones generated in region shown in A; and E and F are digestion of clones generated in region shown in D. Clone numbers 8 and 14 in B and C and clone number 10 in E and F have different size products, but are consistent with the same gene structure inserted in opposite orientation in the vector. All other clones have identical digestion pattern which is consistent with single form of PtKN1 gene.

B - BamHI, N - NcoI, H - HindIII, X - XhoI, A - AclI, E - EcoRI, P - PstI.

2.2.6 Cloning of an *Arabidopsis* KNAT1 mRNA lacking HD Encoding Sequences

Arabidopsis KNAT1 mRNA (NM_116884) has been studied extensively. The absence in any EST database of a KNAT1 variant which lacks HD sequences may reflect the absence of such a splice in this plant, or may indicate very low levels of such an mRNA. I sought to clone an AtKNAT1(hd-) mRNA from *Arabidopsis* embryos. Since the nature, location and magnitude of such a splice was unknown to us, and could conceivably include intronic sequences as I had observed in pine, a PCR strategy using gene specific primers was untenable. However, by making SMART cDNA (Clontech, MountainView, CA) from *Arabidopsis* seeds, 5 days after fertilization and conducting 3' RACE using a primer located 5' to the HD sequence and the adapter primer located at the polyA tail I isolated a cDNA fragment representing a form of KNAT1 mRNA which lacked HD sequences (Figure 2.7, Supplementary Data, Figure A.6, Figure A.7). Subsequently using gene specific primers flanking the splice site I generated several identical cDNA clones, indicating that this appears to be the sole or at least the predominant splice used to remove KNAT1 HD sequences from the primary transcript.

The splice initiates at a GU donor site 59 nucleotides from the normal end of exon 3 and terminates at a UC acceptor site located 167-168 nucleotides within exon 5, thereby excising intron 3, exon 4, intron 4 and the first 168 nucleotides of exon 5. The usage of such non-canonical splice site in *Arabidopsis* has not been observed before. The splice removes the last 29 nucleotides of the KNOX2 domain resulting in loss of the terminal 11 amino acids of the 52 amino acid KNOX2 domain, as well as the entire HD. By contrast, in loblolly pine, the entire KNOX2 domain is present within the PtKN1(hd-) protein.

```

KNAT[ hd- ] -----
NP_192555 MEEYCHDNSTTPQRVSFLYSPI SSSNKNDNTSDTNNNNNNNSSNYGPYNTNNNNHHH
consensus meeqhdnsttpqrvsfl yspi sssnkndnt sdt nnnnnnns snygpynt nnnnhhh

KNAT[ hd- ] -----
NP_192555 GHMLFPHVSSLLPCTTENCFRSDHDCPNNNNNPSVKSEASSRI NHYSMLMRAI HNTGEA
consensus qhrlfphvssl l pqt t encfrsdhdqpn nnnnpsvkseassr i nhysml rr ai hntqea

KNAT[ hd- ] -----
NP_192555 NNNNDNVSDVEAMKAKI I AHPHYSTLLCAYLDCCKI GAPPDVVDRI TAARGDFEARGGR
consensus nnnndnvsdveamkaki i ahphystll qayl dcqki gappdvdr i taarqdf earqqr


KNAT[ hd- ] -----
NP_192555 STPSVSASSRDPELDQFMEAYCDMLVKYREELTRPI QEAMEFI R-
consensus stpsvsassr dPELDQFMEAYCDMLVKYREELTRPI QEAMEFI Rr i esql sml cqspi hi


KNAT[ hd- ] -----
NP_192555 LNNPDGKSDNVGSSDEECEENSGGETELPEI DPRAEDRELKNHLLKKYSGYLSSLKGELS
consensus l nnpdgksd nrgssdeeqennsgget el pei dpr aedr el knhll kky sgl ssl kqel s


KNAT[ hd- ] -----
NP_192555 KKKKKGKLPKEARCKLLTWELHYKWPYPSESEKVALAESTGLDCKGINNFI NGRKRHW
consensus kkkkkgkl pkearqkl l twel hykwypsesekval aestgl dqkqi nnwf i nqrkrhw


KNAT[ hd- ] -----
NP_192555 KPSQDMQFMVMDGLQHPHHAALYMDGHYMGDGPYRLGP
consensus kpsedmqf mvr dgl qhphhaal yrdgHYMGDGPYRLGP


```

Figure 2.7: Alignment of the amino acid sequences deduced from *Arabidopsis* KNAT1(hd-) and KNAT1 (full length) cDNA.

There is 100% identity between the sequences in the overlapping region. Conserved domains are indicated as follows: green - KNOX1; yellow - KNOX2; blue - ELK; red - HD

2.2.7 Identification of a Transcript Lacking HD Sequences in a Monocotyledonous Plant Maize RS1

Magnani and Hake (2008) identified KNATM as a KNOX-like protein lacking HD in dicotyledons, no KNATM sequence could be identified in monocotyledons. I have identified an (hd-) spliced form of RS1 (NM_001111861.1) in maize by BLAST search against the NCBI nucleotide database (Supplementary Data Figure A.8, Figure A.9). This form of mRNA (DY236300.1) lacks part of the ELK domain (the last 19 of the 22 amino acids) and the splice removes the first 27 of the 66 amino acid HD sequences. The key amino acids WW from helix 1 and P from the PYP loops are missing; and the amino acids IN of the conserved WFIN motif from helix 3 are changed to FT.

2.2.8 Expression of *Arabidopsis* KNAT1 mRNA lacking HD Sequences

To assay the relative abundance of mRNAs containing or lacking the HD sequence I conducted semi-quantitative RT-PCR. *Arabidopsis* embryos complete development in about 11 days after which desiccation ensues to produce a fully mature seed after about 14 days [190]. Approximately 5 days after fertilization *Arabidopsis* embryos are entering the ‘torpedo’ stage, a late morphogenic phase of development, and at 10 days they are in maturation phase. Our results show that in 5-day old seeds and in Rosette leaves KNAT1(hd-) mRNA is present in very low amounts relative to KNAT1(HD+) mRNA while in 10-day old seeds and whole plant no KNAT1(hd-) mRNA could be detected (Figure 2.8A). This is in contrast to expression of PtKN1(hd-) mRNA for equivalent stages of development in pine; during late morphogenesis, pine PtKN1(hd-) mRNA is fairly abundant and in late maturing embryos it is essentially the dominant KNAT1 mRNA species (Figure 2.4B).

2.2.9 The KNAT1 and KNATM Genes in *Arabidopsis* Exhibit Different Expression Patterns

During the course of this project, Magnani and Hake (2008) reported a novel member of the *Arabidopsis* KNOX family, KNATM, which is encoded by a small gene on Chromosome 1 (At1g14670). KNATM is expressed, by alternative splicing, via three mRNAs KNATM-A, B and C. The KNATM-B mRNA encodes a full MEINOX domain and in yeast two-hybrid assays was shown to interact with a number of BELL proteins. In experiments where the KNATM-B MEINOX domain was deleted no such interaction was observed and the KNATM-A and KNATM-C proteins, which encode partial MEINOX domains, did not interact with BELL proteins. The KNATM gene was identified in several angiosperm dicotyledonous plants, but no comparable gene was found in monocotyledons. By contrast, here we report that KNAT1-like mRNAs and proteins lacking HD sequences are present in gymnosperms, and both monocotyledonous and dicotyledonous angiosperms. KNAT1-like proteins lacking HD may thus be ubiquitous; the role of KNATM may be complementary to that of KNAT1 in dicotyledons or partially redundant. To examine the relative expression patterns of KNAT1 and KNATM in *Arabidopsis*, we performed semi-quantitative PCR using a primer common to the KNATM mRNAs and a primer specific to the KNATM-B mRNA. Figure 2.8B shows that KNATM-B mRNA is expressed in Rosette Leaves and Seedlings, as was observed by Magnani and Hake (2008), but I detected no expression in seeds either 5 or 10 days after fertilization. These results indicate different patterns of expression for AtKNAT1(hd-) and KNATM-B mRNAs which suggest different roles within development.

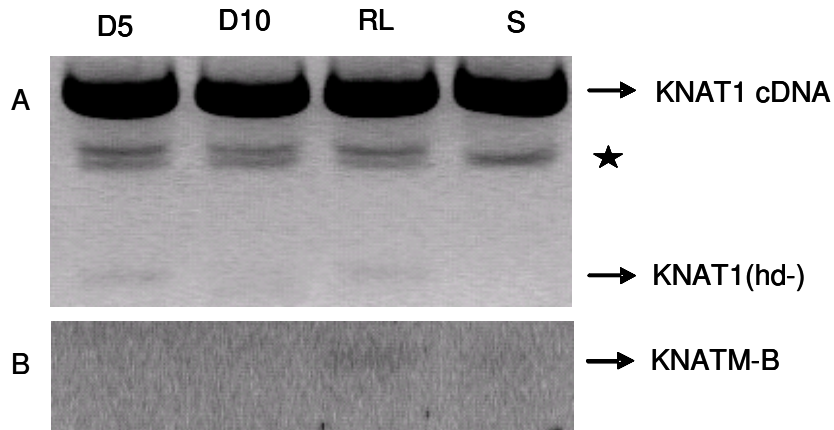
2.2.10 Cloning of *Arabidopsis* KNAT1 mRNA Lacking the MEINOX encoding sequence

Using two gene specific primers located in the 5' and 3' UTRs of KNAT1 mRNA, I isolated a cDNA fragment representing a form of KNAT1 mRNA which lacked MEINOX sequences (Supplementary Data, Figure A.10, Figure A.11). This form of KNAT1 lacks KNOX1 domain and partial KNOX2 domain. The splice initiates at a GU donor site 426 nucleotides from the normal end of exon1 and terminates at the traditional AG acceptor site of intron 2, thereby excising intron 1, exon 2 and intron 2. The splice removes 344 nucleotides starting at position 40 after start codon, the entire KNOX1 domain and the first 45 nucleotides of the KNOX2 domain. The translated product of KNAT1(knox-) mRNA lacks 117 amino acids compared to wild type sequence, starting at position 15 after initiating Methionine, this splice removes the entire KNOX1 domain and first 15 of the 45 amino acids of the KNOX2 domain.

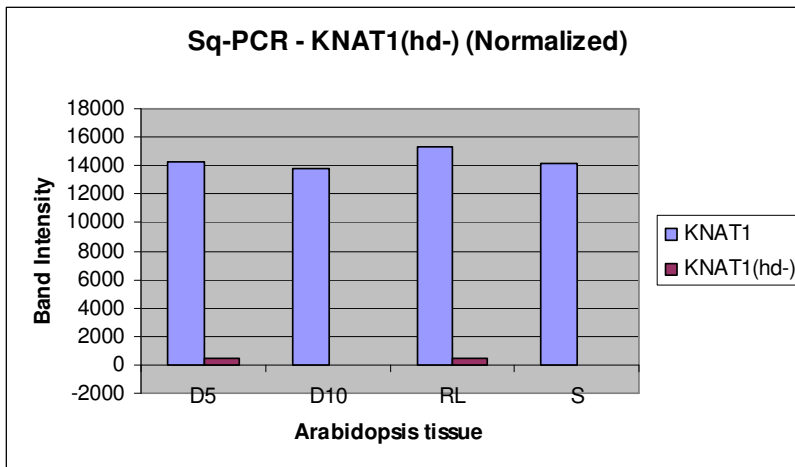
2.2.11 Expression of KNAT1 mRNA lacking MEINOX Sequences

To determine the relative abundance of mRNAs containing or lacking the MEINOX sequence I conducted semi-quantitative RT-PCR. Our results show that in 5 and 10 day old seeds and in seedlings KNAT1(knox-) mRNA is present in low amounts relative to KNAT1(HD+) mRNA and in rosette leaves no KNAT1(knox-) mRNA could be detected (Figure 2.8E).

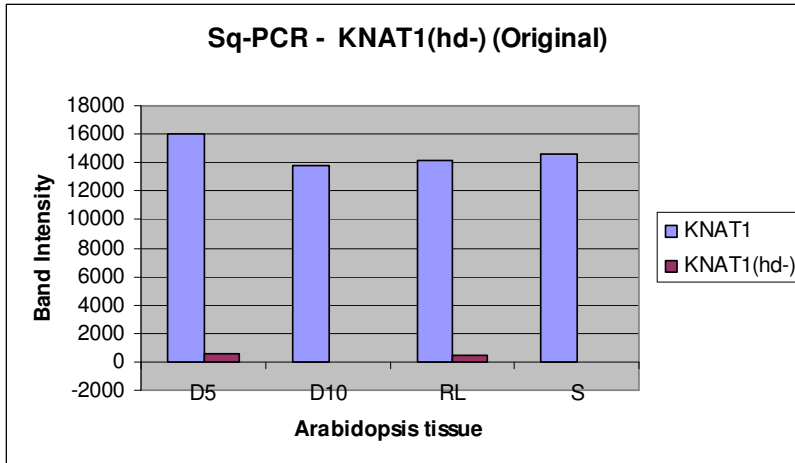
SUMMARY FIGURE SHOWING RELATIVE EXPRESSION OF KNAT1, KNAT1(hd-)



A, B - Enzyme - Advantage2 Taq Polymerase
 No. of cycles - 50
 Volume loaded on gel – 20 ul

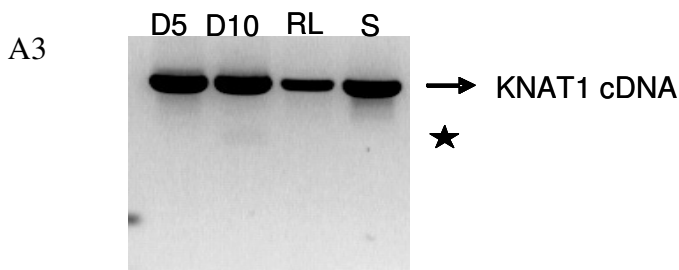


A1 – Normalized graph of sq-pcr for KNAT1 and KNAT1(hd-)

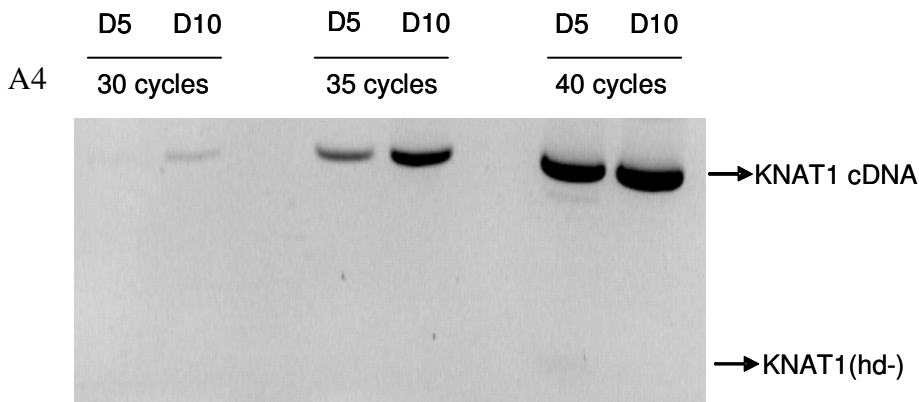


A2 – Original graph of sq-pcr for KNAT1 and KNAT1(hd-)

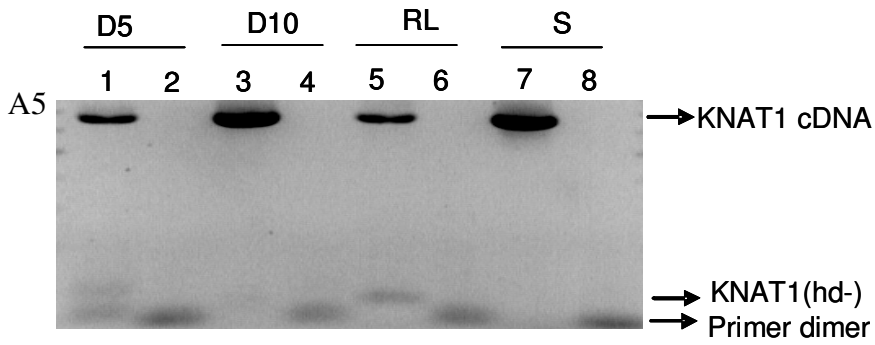
INDEPENDENT spPCR EXPERIMENTS for KNAT1(hd-)



A3- Enzyme - Advantage2 Taq Polymerase
 No. of cycles - 40
 Volume loaded on gel – 10 ul



A4 - Enzyme - Advantage2 Taq Polymerase
 No. of cycles – 30, 35, 40
 Volume loaded on gel – 15 ul



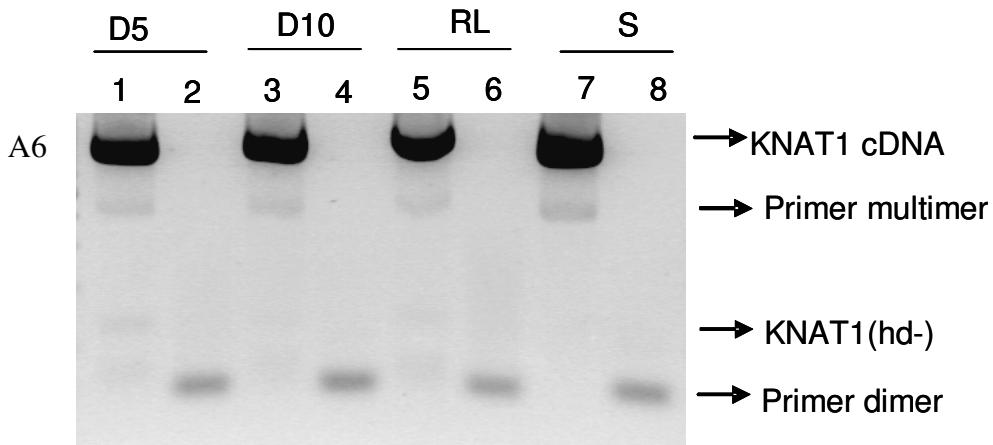
A5 - Enzyme - Advantage2 Taq Polymerase

No. of cycles - 45

Volume loaded on gel - 20 ul

Lanes - 1, 3, 5, 7 are KNAT1

2, 4, 6, 8 are KNATM-B



A6 - Enzyme - Advantage2 Taq Polymerase

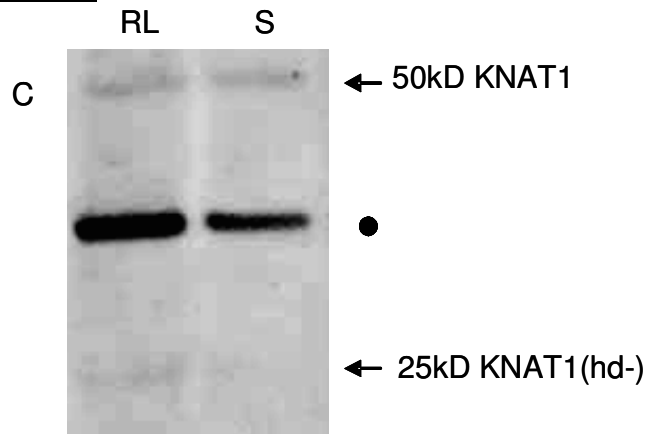
No. of cycles - 50

Volume loaded on gel - 15 ul

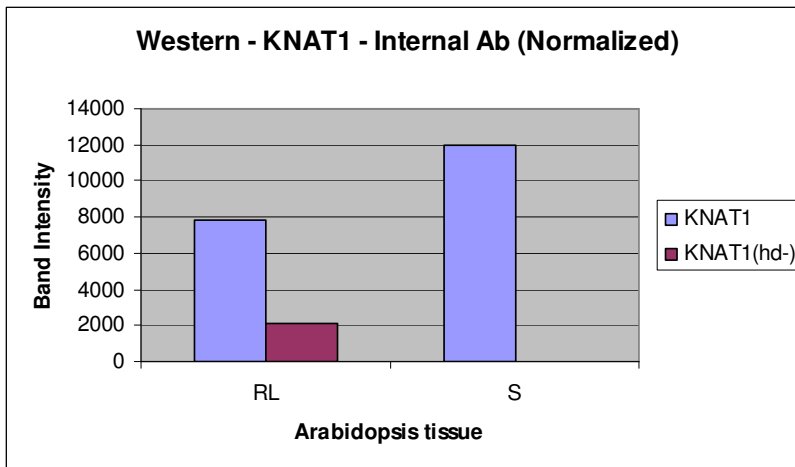
Lanes - 1, 3, 5, 7 are KNAT1

2, 4, 6, 8 are KNATM-B

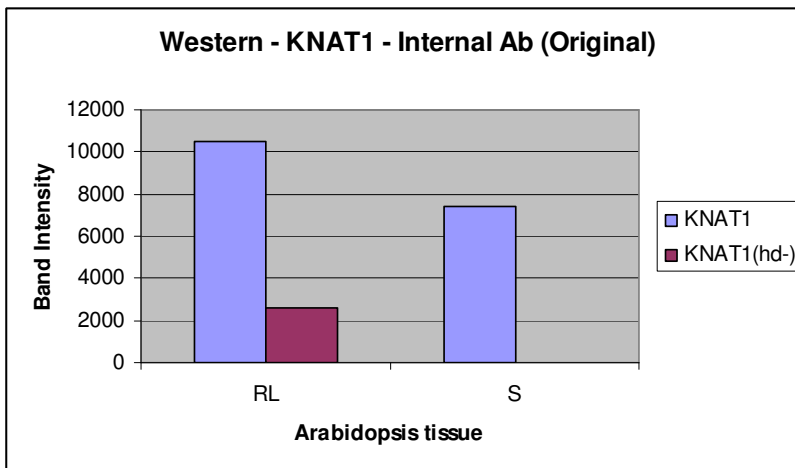
WESTERN BLOTS:



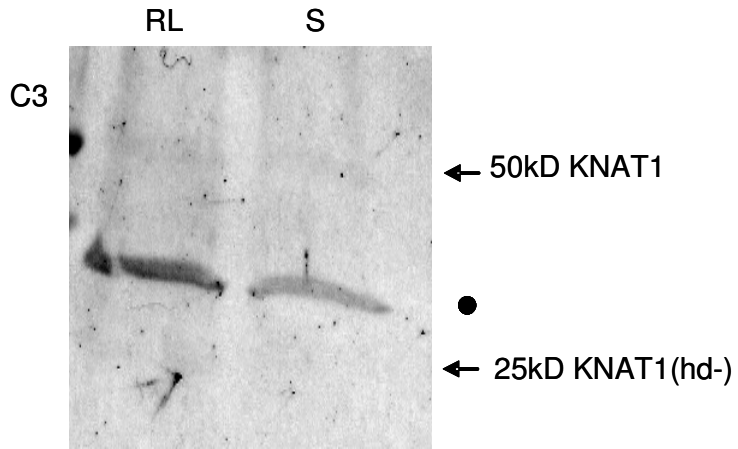
C – Western blotting using Ab that binds to internal region of KNAT1



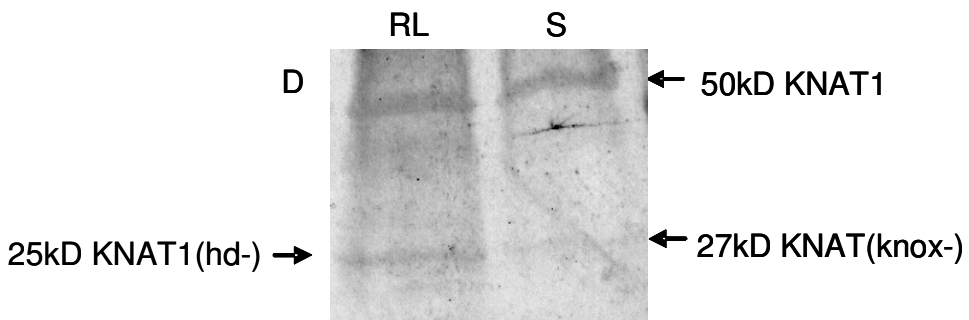
C1 – Normalized graph of western blotting using Ab that binds to internal region of KNAT1



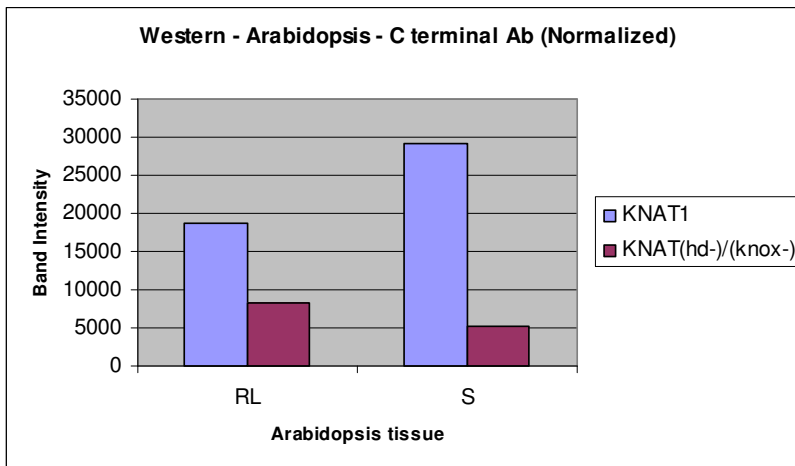
C2 – Original graph of western blotting using Ab that binds to internal region of KNAT1



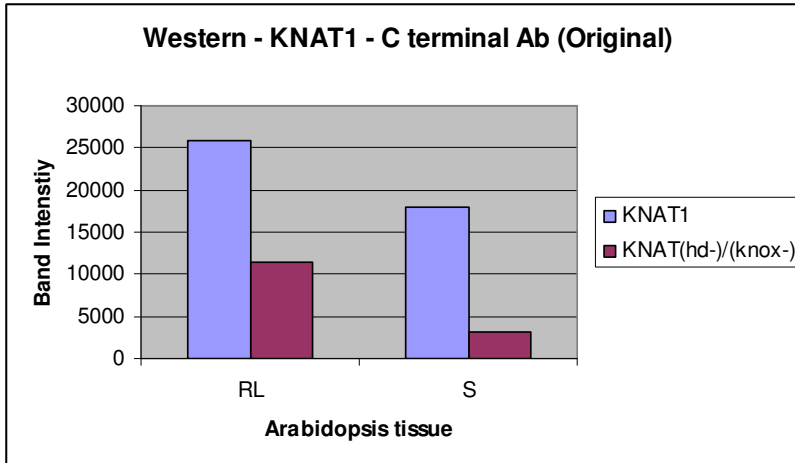
C3 – Independent experiment - western blotting using Ab that binds to internal region of KNAT1



D – Western blotting using Ab that binds to C terminal region of KNAT1



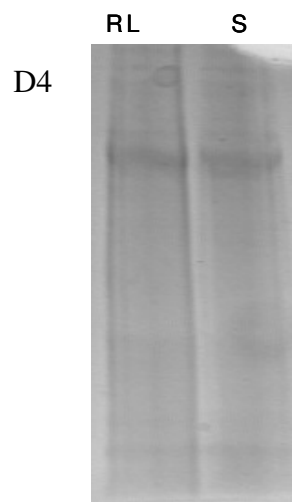
D1 – Normalized graph of western blotting using Ab that binds to C terminal region of KNAT1



D2 - Original graph of western blotting using Ab that binds to C terminal region of KNAT1

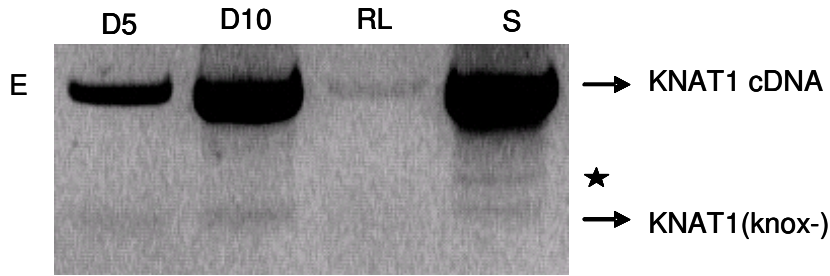


D3 - Independent experiment - western blotting using Ab that binds to C terminal region of KNAT1

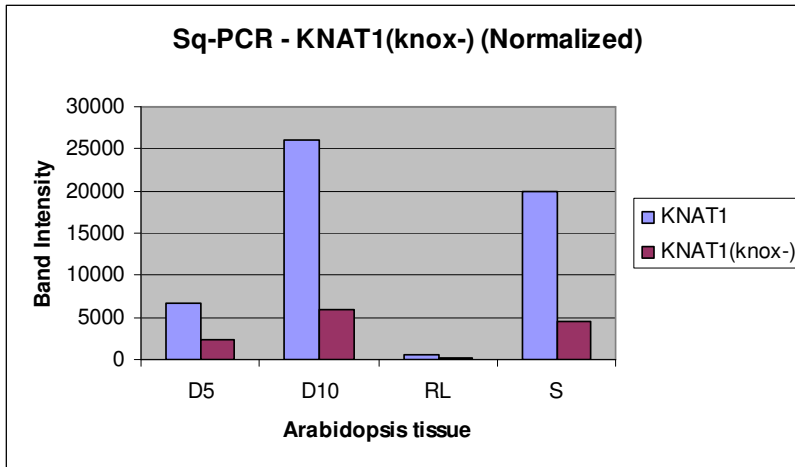


D4 - Control protein loading for Western blotting. 12ug of protein loaded for each lane.

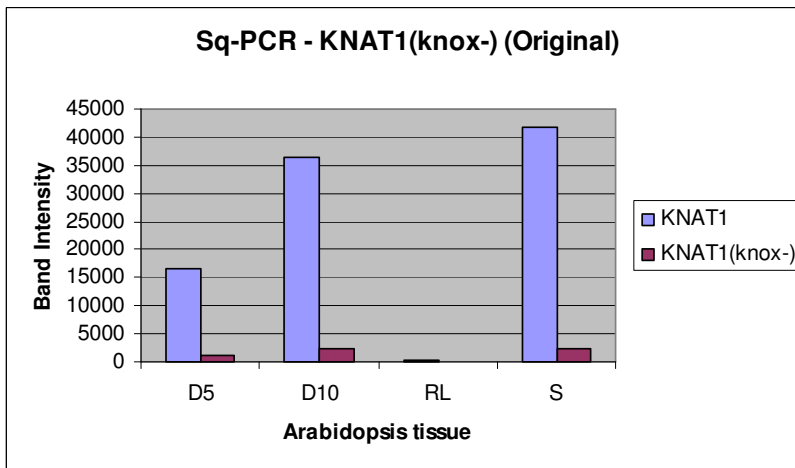
SUMMARY FIGURE SHOWING RELATIVE EXPRESSION OF KNAT1, KNAT1(knox-)



E - Enzyme - Advantage2 Taq Polymerase
 No. of cycles - 40
 Volume loaded on gel – 15 ul

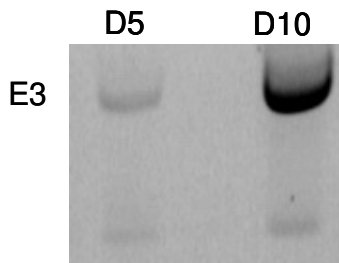


E1 – Normalized graph of sq-pcr for KNAT1 and KNAT1(knox-)



E2 – Original graph of sq-pcr for KNAT1 and KNAT1(knox-)

INDEPENDENT spPCR EXPERIMENT for KNAT1(knox-)



E3 - Enzyme - Advantage2 Taq Polymerase

No. of cycles – 40

Volume loaded on gel – 10 ul

Figure 2.8: Spliced forms in *Arabidopsis* KNAT1 that lack HD or KNOX regions.

(A) RT-PCR analysis of *Arabidopsis* KNAT1(hd-)

(B) RT-PCR analysis of *Arabidopsis* KNATM-B

(A1) Normalized graph of sq-pcr for KNAT1 and KNAT1(hd-)

(A2) Original graph of sq-pcr for KNAT1 and KNAT1(hd-)

(A3) Replicate of Figure2.8A

(A4) Replicate of Figure2.8A

(A5) Replicate of Figure2.8A and Figure 2.8B

(A6) Replicate of Figure2.8A and Figure 2.8B

(C) Western analysis of AtKNAT1 proteins using Antibody sc-19215

KNAT1(hd-) protein is expressed in *Arabidopsis*. Antibody binds to only KNAT1(hd-) and expression is seen in RL. Anti-AtKNAT1 Antibody sc-19215 was purchased from Santa Cruz Biotechnology, CA).

(C1) Normalized graph of western blotting using Ab that binds to internal region of KNAT1

(C2) Original graph of western blotting using Ab that binds to internal region of KNAT1

(C3) Replicate of Figure 2.8C

(D) Western analysis of AtKNAT1 proteins using Antibody sc-19217

KNAT1(knox-) protein is expressed in *Arabidopsis* Antibody binds both KNAT(hd-) and KNAT(knox-) and expression is seen in RL and S respectively. The expected size of KNAT1(hd-) and KNAT1(knox-) from putative protein sequences are 25kD and 27kD. Anti-AtKNAT1 Antibody sc-19217 was purchased from Santa Cruz Biotechnology, CA).

(D1) Normalized graph of western blotting using Ab that binds to C terminal region of KNAT1

(D2) Original graph of western blotting using Ab that binds to C terminal region of KNAT1

(D3) Replicate of Figure 2.8D

(D4) Control protein loading for Figure 2.8C, Figure 2.8C3, Figure 2.8D and Figure 2.8D3

(E) RT-PCR analysis of *Arabidopsis* KNAT1(knox-)

(E1) Normalized graph of sq-pcr for KNAT1 and KNAT1(knox-)

(E2) Original graph of sq-pcr for KNAT1 and KNAT1(knox-)

(E3) Replicate of Figure 2.8E

Day 5 (D5), Day 10 (D10) seeds after fertilization, rosette leaves (RL) and seedling (S) tissue.

★ – Band was cloned and sequenced and shown to be primer multimer (data not shown).

● - Antigen for this antibody is located within the KNOX domain in a region with strong sequence similarity to KNAT2 and KNAT6 proteins which are 35-36kD respectively. Since Antibody sc – 19217 shows greater KNAT1 specificity, and does not detect this band (Figure 6D), the 35kDa species may be KNAT2 and/or KNAT6 proteins.

2.2.12 An OSH1-like Transcript Splice Removes KNOX Sequences and Changes

Reading Frame in Rice

I identified several (knox-) spliced forms of OSH1 in rice, a monocot, by a BLAST search against the NCBI nucleotide database. These transcripts (e.g. AK107637.1, NM_001057674) lack part of the KNOX1 domain (corresponding to the first 19 of the 45 amino acids) (Supplementary Data, Figure A.12). Alignment of AK107637.1 mRNA and full-length mRNA (D16507.1) with the genomic sequence assigns 7 exons to AK107637.1 compared to 5 exons for D16507.1. Translation initiation appears coincident with the major transcript, however, a splice 183 nucleotides after the ATG of AK107637.1 removes 176 nucleotides, creating a 19 nucleotide exon corresponding to the final 19 nucleotides of exon 1 in D16507.1). Thereafter the sequences are identical, and all succeeding splices are common between AK107637 and D16507.1. However the 176 nucleotide splice, effects a change in reading frame and results in a protein lacking MEINOX and HD which has no similarity to OSH1.

2.2.13 Proteins Are Encoded by KNAT(hd-) and KNAT(knox-) mRNAs

To determine whether the KNAT1(hd-) and KNAT1(knox-) mRNAs are translated, I performed Western analysis using protein isolated from rosette leaves and seedling tissue (Figure 2.8D). Using the KNAT1 Antibody sc-19215, which is raised

against a 20 amino acid peptide antigen mapping between aa# 150-200 of NP_192555.1 (Santa Cruz Biotechnology, pers.com. K. Griffin) I detected bands corresponding to the KNAT1 full length protein and the KNAT(hd-) protein in Rosette Leaves (Figure 2.8C). In seedlings I detected the full length KNAT1 cDNA but KNAT(hd-) was absent.. These results are consistent with the RT-PCR results (Figure 2.8A). A very prominent protein band was detected in both samples migrating around 35kD. I have found no AtKNAT1 cDNA capable of encoding a 35kD protein. Since region 150-200 of KNAT1, against which the sc-19217 antibody was raised, encodes the KNOX domain which has very sequence similarity to the KNOX domains of KNAT2 (mw. 35.64 kDa) and KNAT6 (mw. 36.9 kDa) it is possible that the antibody is binding to these proteins (Supplementary Data, Figure A.13).

To analyze KNAT1 expression further, I repeated Western analysis this time using the KNAT1 Antibody sc-19217, which is raised against a 20 amino acid peptide antigen mapping to the C-terminal of KNAT1, between aa# 350-400 of AAM03027.1 (Santa Cruz Biotechnology, pers.comm K. Griffin). This region bears no similarity to other KNAT-like proteins and in this analysis no 35kDa band was seen (Figure 2.8D). The sc-19215 antibody detects full length and KNAT1(hd-) protein in rosette leaves and seedlings. The sc-19217 antibody detects a 25 – 27kDa band in Seedling tissue whereas no band of this size was seen in this tissue using sc-19215. AtKNAT1(knox-) mRNA is present in seedling tissue (Figure 2.8E), and its predicted peptide is 27kD. Since the sc-19217 antibody binds outside the KNOX domain it is likely that the 25 – 27kDa band is the AtKNAT1(knox-) protein. It is notable that while AtKNAT1(HD+) mRNA is more

abundant than the AtKNAT1(hd-) mRNA (Figure 2.8A) the quantities of their proteins don't differ much (Figure 2.8 C, D).

2.3 Discussion

2.3.1 KNOX proteins Containing or Lacking HD Sequences are Produced by Alternative RNA Processing of a Single Gene

Here I report that in loblolly pine embryos two principal forms of KNAT1-like mRNA are produced. One mRNA, PtKN1(HD+), includes the MEINOX and HD sequences found in plant KNAT1-like mRNAs but the second mRNA, PtKN1(hd-), is novel in plants in that while the MEINOX domain is present, the mRNA lacks the ELK and the HD sequence and terminates in a novel 3'UTR. Western analysis reveals that both these mRNAs are translated, the PtKN1(hd-) mRNA and protein are abundant and in late maturing embryos their quantities exceed that of the canonical form. Proteins encoded by PtKN1(hd-) mRNA would be able to form heterodimers but not homodimer since MEINOX domain by itself can form heterodimers but needs the HD for homodimerization [98, 100, 102, 113, 114]. No protein-protein interaction mediated via ELK domain could occur through such a protein and without the HD this protein would not be able to recognize and bind DNA. The PtKN1(hd-) protein may function as a dominant-negative regulator of PtKN1 by binding and sequestering interacting proteins as non-functional dimers. By genomic sequencing I demonstrated that the novel 3'UTR of the PtKN1(hd-) mRNA is encoded by the regions of the gene immediately following the MEINOX domain, a region of the gene that had been designated previously as exon 3. Beyond intron 3 lies the HD sequences encoded by exons 4 and 5. While conifer

genomes are large, contain multi-gene families and many processed pseudogenes, four lines of evidence support the conclusion that PtKN1 gene has a single structure in the pine genome: 1) Twenty nine independently isolated PtKN1 cDNAs were virtually identical in sequence, with a minimum of 0 to a maximum of 8 nucleotide differences over their length. Five PtKN1(hd-) cDNAs were identical to one another and to the overlapping regions of the PtKN1(HD+) cDNAs. 2) Eleven PtKN1(hd-) genomic clones covering exon 1 through the proximal end of intron 3 showed 99.6% identity (1 to 9 nucleotide differences over 2090 bp). The exonic regions of these clones show 100% nucleotide identity with (HD+) and 100% nucleotide identity with (hd-) cDNAs. 3) Restriction enzyme digestion patterns for 14 independent genomic clones of the region from the 5`UTR to exon 2 were identical, similarly 14 independent clones of the region covering exon 4, intron 4 and exon 5 show identical digestion patterns (Figure 2.6). 4) Sixteen independent genomic clones containing approximately 945 bp of the PtKN1 promoter were generated from a region within the open reading frame that is common to both types of transcript. The sequences were identical, containing between 0 and 8 random nucleotide differences over their length. Eight additional independent genomic clones which extended the region to 1670bp beyond the TSS were identical in sequence.

The structure of the gene indicates that the PtKN1(hd-) mRNA could be produced by splicing of the first two introns followed by cleavage-polyadenylation of the primary transcript within intron 3. The PtKN1(HD+) mRNA could be produced by splicing of the first two introns followed by a splice which removes intron 3 and the unique sequences found at the 3'-end of the PtKN1(hd-) mRNA. These processes would be mutually exclusive.

Tian et al. (2007) found that in about 20% of human genes, cleavage-polyadenylation occurred within an intron to produce alternative, truncated mRNA molecules. Weak 5' splice sites and long introns were associated with intron polyadenylation. The choice of alternative polyadenylation site varied with tissue [195], and intronic polyadenylation also varied in different cell lines [196]. A dynamic between splicing and polyadenylation, has recently been shown to control the expression of the thiamin biosynthetic gene, THIC, in *Arabidopsis* [197]. The process is mediated via a riboswitch involving a conserved aptamer sequence which is capable of binding the vitamin B1 derivative, thiamin pyrophosphate (TPP). In the THIC gene, an intron, containing a major 3' end cleavage-polyadenylation site, immediately follows the stop codon. In the absence of TPP, the aptamer, located downstream, can interact with the 5' splice site. This blocks intron removal and the transcript is cleaved to form a strongly translated mRNA bearing a short 3'UTR. When present at high levels, TPP binds to the aptamer, altering its conformation such that 5'splice site interaction does not occur; consequently the intron is spliced out resulting in mRNAs with lengthy 3'UTRs, which are poorly expressed. Riboswitches are well characterized in bacterial systems and have been implicated in controlling expression of several metabolic genes in eukaryotes; however, none of these encode transcription factors.

The KNOX mRNAs have been cloned in Norway spruce and their abundance during embryo development in vitro has been assayed by semi-quantitative PCR. KNAT1 transcripts (HBK3) were expressed at constant level over somatic embryo development. However, these experiments employed primers specific to the 5'- and 3'-UTR of KN1 transcripts which include the HD, thus the expression of the KN1(hd-) mRNA was not

observed. Elevated expression of the HD containing HBK3 protein in somatic embryos of Norway spruce promoted differentiation of immature embryos, increased the frequency with which mature embryos were observed and produced an enlarged stem apical meristem [71]. In cell lines expressing antisense HBK3 constructs (HBK3-A), the differentiation of immature embryos was greatly reduced however immature embryos in sense and antisense lines were not morphologically different. Belmonte et al. (2007) suggested that the antisense construct may be down-regulating an additional KNOX gene, noting the similarity of Norway spruce genes HBK1. Norway spruce embryos do produce KN1 transcripts lacking HD sequences (equivalent to PtKN1(hd-), our unpublished data) thus the phenotype observed could derive from suppressing translation of KN1(hd-) mRNA. When HBK (sense) was expressed in *Arabidopsis* several post embryonic phenotypes were observed such as lobed rosette leaves and dwarf phenotypes, characteristics that were not observed for spruce. Belmonte et al. (2007) suggested that the strength of the different promoters used to express the cDNA may explain the different characteristics. It was observed that AtKNAT-1(hd-) mRNA is expressed at much lower levels than pine PtKN1(hd-) mRNA (Figure 2.4A, Figure 2.8A) thus the relative abundance of (hd-) mRNAs in spruce and *Arabidopsis* may also contribute to the differential phenotypes observed.

2.3.2 The *Arabidopsis* KNAT1 Gene Produces mRNA and Protein Lacking HD

Sequences

I discovered a comparable splicing event that removes the HD sequence from the *Arabidopsis* KNAT1 RNA. The HD spliced KNAT1 mRNA is present at very low levels with respect to the HD-containing KNAT1 mRNA. This is in contrast to the relative

abundance of the PtKN1(hd-) mRNA in pine embryos and the rarity of this spliced form may explain why an *Arabidopsis* KNAT1 mRNA lacking HD sequences has not been reported previously. While the overall quantities of KNAT1(hd-) mRNA may be low, the tissues assayed are complex and this form of mRNA may be very abundant in particular cell types.. Any protein-protein interaction possible through the ELK domain will be lost and without the HD, the KNAT1 (hd-) form will not be able to recognize and bind DNA or localize to the nucleus. This implies that the KNAT1(hd-) protein may function cytoplasmically (because of the lack of NLS which is present at the N-terminal end of the HD), perhaps pairing with proteins thereby preventing their import into the nucleus. The KNAT1 (knox-) proteins lack the C-terminal side of KNOX2 domain. Since the KNOX2 domain is required for dimerization these proteins may be unable to form dimers.

2.3.3 In Monocotyledons KNAT1 like Genes are Spliced to Produce Transcripts lacking HDs

Magnani and Hake (2008) cloned and characterized KNATM, a KNOX-like protein lacking HD sequences. KNATM was found in many dicotyledons, but no homolog could be identified in monocotyledonous plants. By contrast mRNAs of the monocotyledon maize gene rs1 in which the splice has removed some of the HD were identified by search against NCBI database. The protein encoded by this form of RS1 mRNA would be able to form heterodimers, but may not be able to form homodimers. Without the ELK domain and with key regions of HD being either changed or absent, the protein would be unable to bind DNA and so could function as a dominant-negative regulator of RS1 by competing with its interacting partner.

The splice responsible for excising the *Arabidopsis* KNAT1 HD begins in exon3 and removes intron 3, exon 4, intron 4 and part of exon 5. Interestingly, insertion elements which produce dominant mutations of KNAT1 in maize have been shown to locate within exon 3 [198]. Similarly the ‘Hooded’ mutation in barley is located within exon 4. These dominant mutations in non-coding regions allow KNOX expression to expand outside the shoot apical meristems. Greene et al. (1994) suggested that a chromatin organizing element within the intron may have been disrupted by these insertions, resulting in the phenotype [198]. The discovery of a large splice which produces mRNA lacking HD sequences (Figure 2.7, Figure 2.8, Supplementary Data Figure A.6) may provide an alternative explanation. Insertions within introns can reduce splicing efficiency [199, 200]. Insertions within the third and fourth introns of KN1 may affect splicing such that the amount of KN1 (hd-) mRNA and KN1 (hd-) protein is reduced. If these KN1(hd-) proteins do function as suppressors of KN1 activity by interacting with HD-containing proteins, then diminished quantities of KN1(hd-) proteins would result in incomplete suppression of KN1. Spatial expansion of KN1 activity beyond the SAM could be one manifestation of reduction of KN1(hd-) suppressor protein. Such a scenario implies that despite its low abundance, splicing of KN1 to produce KN1(hd-) mRNA is an important regulatory process which is regulated in a cell and tissue-specific fashion.

2.3.4 Splicing Removes MEINOX Sequences from KNAT1 mRNA

I report another spliced form of *Arabidopsis* KNAT1 that lacks the MEINOX encoding sequence. KNAT1(knox-) may not be able dimerize with other protein; but they may be able bind DNA with low affinity as a monomer [201], thus blocking the sites for

KNOX homo or hetero dimers to bind and function. A transcript which lack KNOX1 sequences was also identified for the rice OSH1 gene; splicing creates a frameshift in the mRNA which produces a protein with little similarity to OSH1.

Deletion studies in which domains of KNOX proteins have been removed, exchanged or mutated reveal a variety of phenotypes. Nagasaki et al. (2001) have shown by creating deletion mutants that KNOX1 has suppressive function and KNOX2 is required for dimerization. If the KNOX2 domain or both KNOX1 and KNOX2 domains are deleted from a protein, it cannot form a dimer; but if KNOX1 or KNOX2 or both KNOX1 and KNOX2 domains are deleted the protein can still bind to DNA. Magnani and Hake (2008) have shown that KNATM-B with complete MEINOX domain is able to interact with proteins of BELL family (PNY, BEL1, SAW1, SAW2) and proteins of KNOX family (KNAT1, KNAT3, KNAT4). KNATM isoforms A and C with some MEINOX domain sequence deletion do not interact with either KNOX or BELL family proteins.

Sakamoto et al. (1999) generated chimeric proteins by exchanging different regions of KNOTTED1-like HD proteins in tobacco and they observed that the abnormal phenotypes observed were most severe when the C-terminal half of the ELK domain was exchanged. Nagasaki et al. (2001) constructed several conserved domain deletion constructs and were able to show that the ELK domain acts as a transcriptional repressor of target gene. Based on the phenotype observed, they also suggested that ELK domains may be involved in protein-protein interaction.

Amino acids WW (helix 1), PYP (loop), WFIN (helix 3) are important for DNA binding; in particular PYP and WF are conserved among TALE family proteins in plants

and animals [109]. It has been suggested in animals, the N-terminal end of the HD is required for recognition of the target sequences [74, 109, 110] and the basic amino acid cluster in the N-terminal HD region functions as a NLS [111]. The MEINOX domain by itself can form heterodimers but needs the HD for homodimerization [98, 100, 102, 113, 114]. Spliced mRNAs produced by the human Meis2 gene have MEINOX domains but contain a partial HD deletion. The proteins translated from these mRNAs are unable to bind DNA but compete with the HD-containing Meis2 proteins for dimerizing protein partners, and thus function as a dominant-negative regulator [115]. These results from previous studies help in understanding the function of (hd-) forms.

The naturally-occurring spliced KNAT1 mRNAs which we report encode proteins whose architectures resemble those of the artificial constructs discussed above. Studies in which domains are removed, exchanged or mutated thus predict the range of phenotypes and characteristics that may be expected as these mRNAs are more thoroughly investigated.

KNOX proteins lacking HDs are expressed abundantly in pine embryos but at a very low level in *Arabidopsis* embryos (Figure 2.7). *knox* genes are central to many regulatory processes and *knox* gene misexpression creates pronounced phenotypes [50, 51]. The degree to which the divergent morphology and developmental biology of loblolly pine and *Arabidopsis* embryos can be explained by an alteration in KNOX expression is an intriguing speculation.

2.4 Methods and Materials

2.4.1 Plants and Embryo tissue

Loblolly pine embryos were isolated from pine cones harvested weekly from open-pollinated loblolly pine mother tree 7-56. Cone collection took place from 7/01/04 until 10/15/04 in a commercial pine orchard in Lyons, Georgia, US. Pine seeds were first cracked open to reveal the ovule. The ovules were dissected under microscope to excise the whole embryos which were then flash frozen. Embryos were staged according to the staging system developed by Pullman and Webb (1994). RNA was isolated as described below. *Arabidopsis* tissue and RNA were obtained from Dr. Joe Nairn (UGA) and Dr. John Ohlrogge (Michigan State University).

2.4.2 RNA isolation and cDNA synthesis

Embryos from early stages 1-4, middle stages 5-8, and late stages 9.2-9.3, 9.4-9.5, 9.6-9.7 and 9.8-9.9 were pooled together and total RNA was isolated using TRI-reagent protocol (TR-118, Molecular Research Center, Inc.). Reverse transcription was performed on 500ng RNA from each stages using SMART RACE kit (rapid amplification of cDNA ends from Clontech, MountainView, CA) and the resulting first-strand cDNA containing adaptors ligated to the 3' end was used as template for the PCR reaction. Genomic DNA from loblolly pine tree 7-56 was a gift from Dr. Daniel Peterson, Mississippi Genome Exploration Laboratory (<http://www.mgel.msstate.edu/about.htm>).

2.4.3 Molecular cloning of PtKN1

Using cDNA from late morphogenic stages 5-8 as template, a 500-bp fragment was generated using two primers based on the knotted1 Norway spruce sequence (HBK3,

AF483278). New primers were designed from this sequence and using RACE experiment [202] full length cDNA sequence was obtained. Two gene specific primers 5'-TTGTAGGTCGCCTGGCCAAAGAGCTGA-3' (outer) and 5'-TCACCCAAATGCTGGCTGAATGCTGTGA-3' (nested) were designed in the 5' UTR and were 107-80 and 65-37 bp upstream of the start codon respectively. These outer and nested primers were used with universal primer mix (UPM) and nested universal primer (NUP), respectively, to perform 3' RACE for isolation of full length cDNA. PCR mixture of 50ul volume containing 2.5µM of each dNTPs, 1X PCR buffer, 1 unit of Advantage Taq polymerase, 1µM of the gene specific primer and 1µM of the universal primer mix for the outer reaction or 1µM of the nested universal primer for the nested reaction were used. 3' RACE reaction was performed with conditions of 94° for 2 minutes; 40 cycles of 94° for 30 seconds, Tm for 30 seconds and 72° for 2 minute; followed by 72° for 10 minutes. cDNAs of various size were isolated by 3' RACE. The PCRs were performed in Eppendorf thermocycler (Westbury, NY).

2.4.4 Molecular cloning of KNAT1 spliced forms lacking HD and MEINOX domain encoding sequences

Two gene-specific primers 5'-ACGGCGGCACGGCAAGACTTTGAGGCT-3' (outer) and 5'-CCCGGAGTTAGATCAATTCATGGAAGCATAC-3' (nested) were designed to an internal region (KNOX domain coding sequence) based on KNAT1 cDNA and were 501-528 and 572-603 bp downstream of the start codon, respectively. These outer and nested primers were used with universal primer mix (UPM) and nested universal primer (NUP) respectively to perform 3' RACE experiment [202] with cDNA from *Arabidopsis* seeds 5 days after fertilization. Partial cDNA sequence of KNAT1 was

cloned from which C-terminal sequence of KNOX2, and the entire ELK and HD sequences were spliced out. PCR mixture of 50ul volume containing 2.5μM of each dNTPs, 1X PCR buffer, 1 unit of Advantage Taq polymerase, 1μM of the gene specific primer and 1μM of the universal primer mix for the outer reaction or 1μM of the nested universal primer for the nested reaction were used. 3` RACE reaction was performed with conditions of 94° for 2 minutes; 40 cycles of 94° for 30 seconds, Tm for 30 seconds and 72° for 2 minute; followed by 72° for 10 minutes.

Two gene-specific primers 5`-GAGTCGTCTAGTCGTCTGGATTTGATGTGG-3` (forward) in the 5`UTR and 5`-TCGTTCGGGGAGTTTGAGACTTCCTAGTCA-3` (reverse) in the 3`UTR, were designed and were 140-170 and 133-163 bp upstream of the start codon and downstream of the stop codon, respectively. Using cDNA from the same tissue, PCR was performed and a KNAT1 sequence was cloned that lacks the entire KNOX1 region and partial KNOX2 region. PCR mixture of 50ul volume containing 2.5μM of each dNTPs, 1X PCR buffer, 1 unit of Advantage Taq polymerase, 1μM of the gene specific primer were used. PCR reaction was performed with conditions of 94° for 2 minutes; 40 cycles of 94° for 30 seconds, Tm for 30 seconds and 72° for 2 minutes; followed by 72° for 10 minutes. The PCRs were performed in Eppendorf thermocycler (Westbury, NY).

2.4.5 Semi-quantitative PCR

Pine

Using cDNA from different pine embryo stages as template, RT-PCR was performed with two gene-specific primers for PtKN1(HD+), PtKN1(hd-) and albumin as control. RNA extraction and cDNA preparation are as mentioned above. Primers used

for the analysis were 5`- AAGACGGAGGGAGGTGGATCTTCGGAGG -3` and 5`- GTTTTACGCCGAAGGTTCTTTCTAGCAAT-3` located 895 and 1305 bp downstream of start codon respectively for PtKN1(+); 5`-AGGGCCATGGACGATATCGAGGGAT-3` and 5`-ACGAACACTAACCTCCCAGCGGTTT-3` located 1081 and 1302 bp downstream of the start codon respectively for PtKN1(-) and 5`- AGAGATTGTCTGCTTGTCGGGACT -3` and 5`- TTCTTCTTCTTCCCTTCCGCGCC-3` located 166 and 376 bp downstream of the start codon respectively for albumin. PCR mixture of 50ul volume was prepared as mentioned above. The cycling conditions were 94° for 2 minutes; 30 cycles for PtKN1(hd-) or 35 cycles for PtKN1(HD+) and albumin of 94° for 30 seconds, Tm for 30 seconds and 72° for 1 minute; followed by 72° for 10 minutes.

Arabidopsis

For *Arabidopsis* KNAT1 and its spliced forms, and KNATM-B, RT-PCR was performed using cDNA from Day 5 and 10 dpa seeds, rosette leaves and seedlings as template. RNA extraction and cDNA preparation are as mentioned above. Primers used for the analysis were 5`- CCCGGAGTTAGATCAATTCATGGAAGCATAC-3` and 5`- AGAAAGCAACGAGAGGTTGTTATTATGTGGC-3` located 622-591 bp upstream of stop codon and 20-51 bp downstream of start codon respectively for KNAT1(hd-); 5`- GAGTCGTCTAGTCGTCTGGATTTGATGTGG-3` and 5`- AGCCACATAATAACAACCTCTCGTTGCTTTCTTGT-3` located 140-170 bp upstream of the start codon and 19-54 bp downstream of stop codon respectively for KNAT1(knox-); and 5`- TGGATGTGAAGAAAGATGAGAAC-3` and 5`- TGACTCGAAGCGTATTGC-3` located at 2nd position of start codon and 100-82 bp

upstream of stop codon, respectively, for KNATM-B. PCR mixture of 50ul volume was prepared as mentioned above. The cycling conditions were 94° for 2 minutes; 40 cycles for KNAT1(knox-) or 50 cycles for KNAT1(hd-) of 94° for 30 seconds, Tm for 30 seconds and 72° for 0.5 - 1 minute; followed by 72° for 10 minutes.

2.4.6 Western blot analysis

Pine

Embryos from early stages 1-4, middle stages 5-8, late stages 9.2-9.3, 9.4-9.5, 9.6-9.7 and 9.8-9.9 were pooled together, homogenized and sample buffer containing 10% glycerol, 2% SDS, 5% b-mercaptoethanol was added (1ml sample buffer for 0.2 g tissue). The samples were boiled for 5 minutes at 95° and then centrifuged at 4° for 5 minutes at 14000 rpm. The supernatant was used as protein sample for western blotting. 20ug of each sample was loaded on 4-20% SDS gel from Invitrogen (Carlsbad, CA) and run at 200V for an hour. Transfer was done to nitrocellulose membrane overnight at 20V. The membrane was then blotted with blocking solution (5% milk), primary antibody and secondary antibody for an hour with two rinses and 3 washes of 15, 5 and 5 minutes in between. Antibody synthesized using peptide in the N-terminal region of the protein sequence that is common to PtKN1(HD+) and PtKN1(hd-) forms was used. The antibody was synthesized by Genscript Corporation (Piscataway, NJ) and the peptide sequence used to generate them is 15 amino acids in length. Goat anti-rabbit IgG was used as secondary antibody. Bands were detected using NBT-BCIP color-substrate detection and sizes were determined using Kaleidoscope precision plus protein marker from Biorad (Hercules, CA).

Arabidopsis

Western blotting for KNAT1 and its spliced forms was performed using commercial polyclonal antibody (Santa Cruz biotechnology, Santa Cruz, CA). KNAT1 Antibody sc-19217, is raised against a 20 amino acid peptide antigen mapping between aa# 350-400 of AAM03027.1 (Santa Cruz Biotechnology, pers. comm K. Griffin). KNAT1 Antibody sc-19215, which is raised against a 20 amino acid peptide antigen mapping to the C-terminal of KNAT1, between aa# 150-200 of NP_192555.1 (Santa Cruz Biotechnology, pers.comm K. Griffin). The epitope was raised against KNAT1 protein that binds to the internal sequence. Total protein from *Arabidopsis* rosette leaves and seedlings was isolated, run on SDS-page and analyzed by Western blotting the using same protocol.

2.4.7 Quantification of sq-pcr and western blotting images

Equal amounts of PCR products were separated by gel electrophoresis and Ethidium Bromide-stained bands were visualized. Images were collected by a Monochrome CCD Camera (Hitachi KP-M1U). The digital images of gels for spPCR and Western blots for proteins were quantified using ImageJ software (<http://rsbweb.nih.gov/ij/>) from NIH. The region of the band was selected across all the lanes and was plotted. The peaks for each lane were measured and the numbers were used to make bar charts in Microsoft excel. The background value was subtracted from the band intensity value to get an accurate reading. For Normalization, the signal intensity of albumin band at each stage was used; the intensity of the albumin band in one stage was selected as the norm and the values for albumin for other stages were brought to same level ('100%'). The multiplication factor needed to bring albumin for that stage up to 100% was then used to multiply the values for HD+ and hd- forms for that stage. The

normalized numbers for all the signals were then plotted. For western blotting, a blank/smear region of the image (where there are no bands) was used as control and same procedure was used for normalization.

2.4.8 DNA purification, Cloning and Sequencing

The PCR products were run on 1.5% agarose gel to determine their size and were purified using QIAquick kit from Qiagen (Valencia, CA) and the resulting products were cloned into a pGEM-T Easy vector system from Promega (Madison, WI). The product sizes were confirmed by running colony PCR. Selected colonies were grown in bacterial culture and plasmids were purified using the Wizard SV minprep kit from Promega (Madison, WI), sequenced and analyzed.

2.4.9 Cloning of the Knotted1 gene

7.7 kb genomic sequence of PtKN1(HD+/hd-) gene was isolated by using a combination of PCR and genomic walk and 50 kb genomic DNA which was obtained from Dr. Daniel Peterson (Mississippi State University). Primers used and their positions are shown in the supplementary data table A.1 and Figure A.5. Each PCR mixture of 50ul volume contained 2.5μM of each dNTPs, 2μM of each primer, 1X PCR buffer and 1 unit of Advantage cDNA PCR mix and was performed in the Eppendorf thermocycler. PCR conditions used were 94° for 2 minutes, followed by 35-40 cycles of 95° for 30 seconds, Tm° for 30 seconds and 68° for 2-5 minutes, followed by 68° for 10 minutes.

2.4.10 Alignments

Clustal W and BOXSHADE programs from Biology Workbench were used for alignment and viewing. (<http://workbench.sdsc.edu/>). For all the supplementary figure

alignments (shown in Appendix A), the nucleotide alignments is not shown for whole sequence, are shown only for significant regions covering the splice.

2.4.11 Accession Numbers

The GenBank accession numbers of sequences referred to or used in alignments in this study are: *Arabidopsis* KNAT1 mRNA - NM_116884 and KNAT1 gene - NC_003075.4, picea abies HBK3 -AF483278, maize RS1 - NM_001111861.1, RS1 splice form that lacks partial ELK domain and the HD - DY236300.1, rice OSH1 mRNA - D16507.1, OSH1 spliced forms that lack partial KNOX1 domain - AK107637.1 and NM_001057674. KNATM-B locus Id is At1g14670.

CHAPTER 3

PtKN1 PROMOTER ANALYSIS

3.1 Introduction

A suite of genes is required for correct embryo development, but alteration in a single gene can have profound effects on development. Very different morphologies, life cycles and characteristics of plants can be explained by the difference in expression of regulatory genes which are mostly caused by mutations in promoters and enhancers [184]. Various mutational studies have shown that plant class I KNOX genes play an important role in regulating meristem function. Their loss or gain of function mutation can affect overall plant height, leaf shape, meristem development and floral development. [44, 48, 53-56, 59-68]. And therefore KNOX genes are important and could provide clues to embryo development in gymnosperms and angiosperms.

I have cloned cDNAs corresponding to a homeobox gene, knotted1 from Loblolly pine (referred to as PtKN1(HD+)) and in addition have found a novel form which lacks important domains ELK and HD (referred to as PtKN1(hd-)) (Chapter 2, Sheth and Cairney, submitted). I have shown that both forms are product of a single PtKN1 gene and I have shown the expression patterns of these cDNAs and their proteins over the course of pine embryo development. Some questions important to the understanding of the functions of PtKN1 are, when and where is it expressed, how is its expression regulated i.e. what transcription factors bind to regulate its activity? KNAT1 is a single gene in *Arabidopsis* but multiple gene copies are common in pine. Are all PtKN1 genes regulated in identical fashion or are some copies only expressed in certain tissue or at

specific stages of development. Isolation of promoter and identification of regulatory elements is a step towards answering these questions. In this study, I have isolated its promoter region by genome-walking technique, (Clontech, Palo Alto, CA) and identified the 5' end of the transcript using 5' RACE technique. I have compared PtKN1 promoter sequence to promoters of KNAT1-like genes from other species; by a method called phylogenetic footprinting, which is comparison of conserved regions among orthologous regulatory sequence that are thought to be of functional importance [203-206]. The presence or absence of cis-regulatory elements in PtKN1 promoter and its orthologous sequences, their position and number of occurrence was determined in an attempt to understand PtKN1 regulation and function.

3.2 Materials and Methods

3.2.1 Cloning of the Knotted1 promoter

To clone the promoter region of PtKN1 gene, the genome walker kit from Clontech (Palo Alto, CA) was used and 5' genomic walk was performed. Four genomic libraries were made from pine genomic DNA that was cut using four different restriction enzymes EcoRV, DraI, PvuII and SspI and then ligated to adaptor sequences. Two gene specific primers 5'-TTGATCGAGTTCTGGGTCCATTCC-3' (outer) and 5'-GCCAAGACGAGGCTTGTGCAGCTGCTGT-3' (nested) were designed in the ORF and were located at 672-696 bp and 15-43 bp downstream of the start codon. Two rounds of PCRs were performed using adaptor primers (AP1 or AP2) which bind to the adaptor sequences with outer or inner gene specific primers, respectively. The genomic libraries were used as template for 1st PCR and the products obtained were used for nested PCR. Each PCR mixture of 50ul volume contained 2.5µM of each dNTPs, 1µM of each primer,

1X PCR buffer and 1 unit of Advantage cDNA PCR mix. PCR conditions used were 7 cycles (for 1st PCR) or 5 cycles (for 2nd PCR) of 94° for 25 seconds followed by 72° for 4 minutes and 32 cycles (for 1st PCR) or 22 cycles (for 2nd PCR) of 94° for 25 seconds followed by Tm for 4 minutes and additional 67° for 4 minutes after the final cycle. The PCRs were performed in Eppendorf thermocycler. Once the product was sequenced new primers were designed to walk further upstream to get more promoter sequence and using same PCR reaction mix and conditions more sequence was obtained. The sequence of the new primers are 5`- AGCCAAATAGTTAGTGTGACGTTGGT -3` (outer) and 5`- GAAATCACTTCTCCTTCTTCTTCCA -3` (nested) and they were 197bp and 35bp from the end of the 5` upstream sequence obtained from the 1st walk.

3.2.2 5` RACE

Two gene-specific primers 5`- ACTCCCTATCCTCTACTGCATGGTGATCCACTTCCTG -3` (outer) and 5`- TCACCACCACTGCCGTTCTCGACTTCCTCCGA -3` (nested) that are specific to PtKN1(HD+) and 5`- GCGATGAATACGAACACTAACCTCCCAGCGGTTTCCTG -3` (outer) and 5`- CCTGTAAGCCCCTTCGACCTGGATATTTAGTACTGGG -3` (nested) that are specific to PtKN1(hd-) were designed. The primers in order were 388-420bp and 341-378 bp upstream of the stop codon and 343-381bp, 347-310bp downstream of the stop codon, respectively. These outer and nested primers were used with universal primer mix (UPM) and nested universal primer (NUP), respectively, to perform 5`RACE to determine the TSS of the PtKN1 cDNAs. PCR mixture of 50ul volume containing 2.5µM of each dNTPs, 1X PCR buffer, 1 unit of Advantage Taq polymerase, 1µM of the gene-specific primer and 1 µM of the universal primer mix for

the outer reaction or 1 μ M of the nested universal primer for the nested reaction were used. 3' RACE reaction was performed with conditions of 94° for 2 minutes; 40 cycles of 94° for 30 seconds, Tm for 30 seconds and 72° for 2 minutes; followed by 72° for 10 minutes. The PCRs were performed in an Eppendorf thermocycler.

3.2.3 DNA purification, Cloning and Sequencing

The PCR products were run on 1.5% agarose gel to determine their size and were purified using QIAquick kit from Qiagen (Valencia, CA) and the resulting products were cloned into a pGEM-T Easy vector system from Promega (Madison, WI). The product sizes were confirmed by running colony PCR. Selected colonies were grown in bacterial culture and plasmids were purified using the Wizard SV minprep kit from Promega (Madison, WI), sequenced and analyzed.

3.2.4 KNAT1-like promoter orthologs

KNAT1-like orthologous promoter sequences were obtained from the NCBI database. Upstream of the translational start site, 1488 bp of *Picea glauca* KN2 promoter sequence [96], 2006bp of *Arabidopsis thaliana* KNAT1 promoter sequence [60], 2143 bp of *Cardamine hirsuita* BP promoter sequence [207], 1820bp of *Zea mays* KN1 promoter sequence [208] and 1828bp of *Triticum aestivum* wknox1d promoter sequence [209] were used.

3.2.5 Comparison of PtKN1 promoter to orthologous KNAT1-like promoters

Since the TSS for PtKN1(hd-) was very close to the translational start site (see section 3.3.2) and the TSS for all promoters was not known the comparison was made from the translation start site. Promoters were analyzed using FOOTPRINTER [210]

(<http://genome.cs.mcgill.ca/cgi-bin/FootPrinter3.0/FootPrinterInput2.pl>). This program takes multiple sequences as input and looks for regions of functional conservation amongst them based on phylogenetic relationship. It conducts a conserved motif search that can be 6-12 bp long, and allows for 0-6 bp mutation in the motif. It also allows for motif loss and also inversion. This gives liberty of finding motifs that are only present in common amongst all sequences or are species specific but not all thus allow determining the evolutionary distance among the KNAT1-like promoters. I searched for motif size that were 10 bp long allowing 2 mutations in the motifs.

3.2.6 Identification of regulatory motifs from known database

With the help of a motif comparison program called STAMP, the conserved motifs identified by footprinter were searched against PLACE database, a database of known regulatory motifs [211, 212] (<http://www.benoslab.pitt.edu/stamp/>).

3.2.7 Regulatory motif search using known database

Putative cis-regulatory elements were identified in the PtKN1 promoter using Plant-CARE (Plant cis-acting regulatory elements) [213] (<http://sphinx.rug.ac.be:8080/PlantCARE/index.html>) and PLACE (A database of Plant cis-acting regulatory DNA elements) [214] (<http://www.dna.affrc.go.jp/htdocs/PLACE>) databases. A selected number of motifs found from these databases were used for comparison of PtKN1 promoter to its orthologous promoters as mentioned above. The presence or absence of these motifs in other KNAT1-like promoters, their position and number of occurrence was determined.

3.3 Results and Discussion

3.3.1 Cloning of PtKN1 gene promoter

I have isolated and characterized Knotted1 transcription factor PtKN1(HD+) and its HD-lacking spliced form, PtKN1(hd-) in loblolly pine (Chapter 2, Sheth and Cairney, submitted). Some questions important to the understanding of the functions of PtKN1 are, when and where is it expressed, how is its expression regulated i.e. what transcription factors bind to regulate its activity? Isolation of promoter and identification regulatory elements is a step towards answering these questions. I employed the genome walk technique to isolate the promoter sequence for PtKN1 gene using primers in the ORF (Figure 3.1A) in combination with an adapter primer (AP1 or AP2) provided with the kit. Sixteen individual clones about ~1000bp in length were isolated (Figure 3.1B). All 16 clones were nearly identical in sequence. Taking one clone as a reference we observed between 0 and 8 nucleotide differences over 1.0kb. These differences are random, most likely due to sequencing error. New primers were designed at the end of this clone to walk further upstream to get more sequence (Figure 3.1A). Eight individual overlapping clones about ~800bp in length were isolated (Figure 3.1C). After making contig of the sequences from both walks, a total of 1730 bp sequence upstream of the start codon was obtained. The sequence of eight independent genomic clones representing sequence 1236 – 1669 bp upstream of the putative TSS (see section 3.3.2) were nearly identical. The lack of sequence heterogeneity within 1,730 bp of PtKN1 promoter sequence is consistent with a single form of PtKN1 gene.

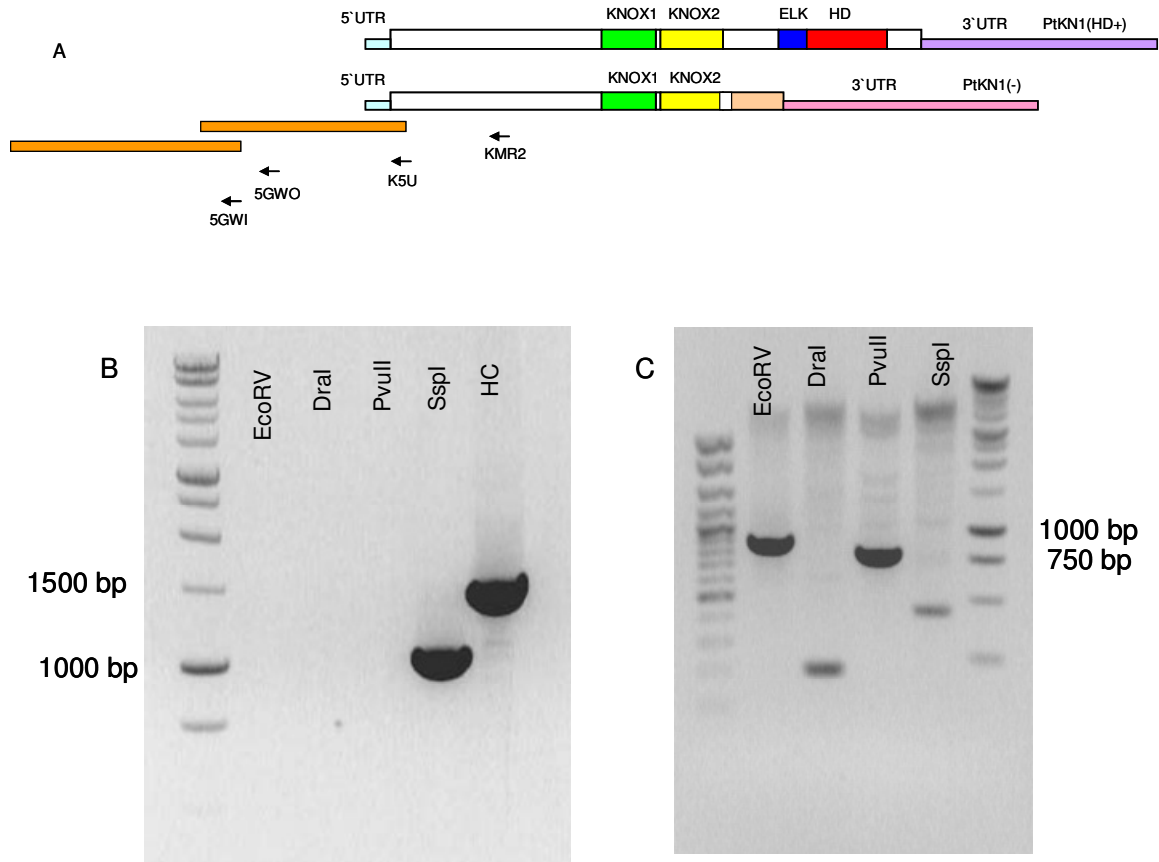


Figure 3.1: Promoter isolation by genomic walk.

(A) Primers locations for promoter isolation. KMR2 and K5U were used for first walk which bind to both PtKN1(HD+) and PtKN1(hd-) transcripts. 5GWO and 5GWI were designed at the end of new sequence obtained from 1st walk to perform 2nd walk. PCRs were performed using genomic libraries digested with enzymes EcoRV, DraI, PvuII and SspI and products were run on gel.

(B) Gel picture for 1st genomic walk. Products of SspI library were cloned and sequenced. (HC is the human control). 16 clones with nearly identical 1000bp sequence were obtained.

(C) Gel picture for 2nd genomic walk. Products of EcoRV and PvuII libraries were cloned and sequenced. 8 clones with nearly identical 800bp sequence were obtained. After making contig of the sequences from both walks, a total of 1730bp sequence was obtained upstream of the start codon.

3.3.2 Identification of transcription start site

The TSS was determined using 5' RACE. Reverse primers that will bind only to PtKN1(HD+) or PtKN1(hd-) form were designed based on the sequence of their cDNAs and were used in combination with two adapter primers (UPM – outer PCR and NUP – nested PCR) provided with the kit (Figure 3.2A). The products of outer PCR were used as template for nested PCR. Since the products from 1st PCR were cleaner they were cloned and sequenced. Multiple individual clones were sequenced for each form to confirm the end of the transcript. Two different start sites were identified for PtKN1(HD+) and a single start site was identified for PtKN1(hd-). The TSS of 3 independent PtKN1(HD+) clones was 494 bp upstream of the translational start site and of 2 independent clones was 343 bp upstream of the translational start site. The TSS of 4 independent PtKN1(hd-) clones was 61 bp upstream of the translational start site. The positions of the transcriptional start site relative to the translational start site are shown on a bar which represents the 1730bp promoter (Figure 3.2B). A TATA box at position -23 and a CAAT box at -114 were found for PtKN1(HD+) with TSS at 494 bp upstream of start. No TATA box or CAAT box were found for PtKN1(HD+) with TSS at 343 bp upstream of the start codon suggesting that like the TATA box lacking promoters, it may utilize the initiator element to initiate transcription from this site. Alternatively the TSS at position 494 would be the preferred start for PtKN1(HD+) transcript. A TATA box at position -35 and a CAAT box at -129 were found for PtKN1(hd-) with TSS at 61 bp upstream of start codon.

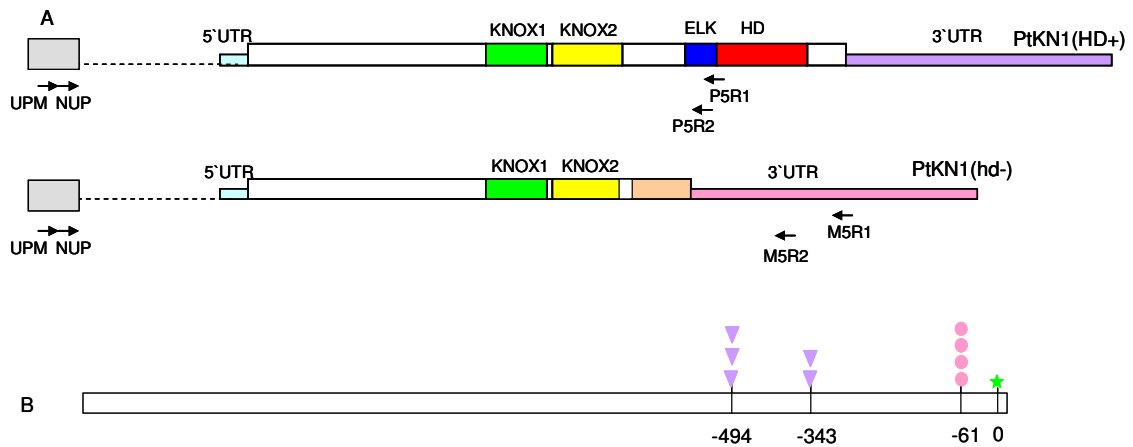


Figure 3.2: Determine the transcription start site of PtKN1 transcripts.

(A) 5` RACE was used to determine the start site. Location of gene-specific primers that bind to either PtKN1(HD+) or PtKN1(hd-) sequences are shown in 2A. In combination with adaptor primers UPM and NUP which are present at the end of transcripts, 5` RACE was performed to determine the end of HD+ and hd- transcripts.

(B) Bar represents the 1730 bp PtKN1 promoter sequence. Green star represents the translational start site which is same for both the transcripts and is considered as position 0. A single TSS was isolated for hd- transcript and two different sites were isolated for HD+ transcript. The pink circle represents the TSS for PtKN1(hd-) transcript which is located 61 bp upstream of the start codon; and the purple inverted triangle represents the TSS for PtKN1(HD+) transcript which are located 343 and 494 bp upstream of the start site. The numbers below the bar represent their positions relative to the start codon. TATA and CAAT boxes were identified for the hd- TSS and for the HD+ TSS at position 494, but not for the HD+ TSS at position 343. Number of triangles and circles represent number of independent clones for that particular start site.

Recent studies on some mouse and human promoters have shown that the TSS within core promoter varies among different tissues. This was explained to be due to differential CpG methylation, promoter structure and imprinting that affect the transcription machinery [215, 216]. Since the TSSs for PtKN1(HD+) and PtKN1(hd-) are different cell or tissue specific gene regulation is a possibility. Alternative start site usage has been shown to produce alternatively spliced products [216]. The human c-myc

gene has two different promoters and transcripts originating from these have different termination sites [217]. Likewise, the human MID1 gene employs multiple promoters and multiple polyadenylation sites. The point at which transcripts are cleaved and polyadenylated seems to be determined by which promoter was used [218]. The Human Pur family proteins are involved in both DNA and RNA metabolism. The PURG gene is normally expressed via a single exon gene from multiple promoters producing the PURG-A mRNA with a 5'UTR of up to 1200 nt, an open reading frame of around 1000 nt and a 3'UTR of about 900bp. A novel mRNA, PURG-B was recently isolated that is produced by read-through of the termination / polyadenylation signal to produce a very lengthy transcript from which a 30kb intron is then removed to produce a novel PURG protein [219]. The mechanism has not been elucidated but the parallels with PtKN1 are clear.

The production of mRNAs which have extended 5'UTRs could, in theory, be achieved by using multiple promoters and a single cleavage/termination site. However it has been observed that when multiple promoters are employed, termination occurs at several sites and that initiation of transcription at one particular site will always be followed by termination at only one of these sites. This indicates some level of regulation proceeding from the chosen initiation site which dictates the position at which the transcript will terminate. A further level of complexity occurs with PtKN1 where read-through of a transcription termination signal and additional splicing of one of the mRNA are associated with the choice of promoter. The 5'UTR of PtKN1 contains binding sites for three different SM protein/snRNP proteins involved in RNA processing (Fig 3.3A) [220]. Commencing transcription at 494, 343 or 61 nucleotides upstream of the

translational start site produces transcripts which contain different numbers of these motifs. The number of mRNAs ending at each site, while a small sample, suggest that the 494 and 61 transcripts are most abundant. All the SM protein binding sites would be present in the transcript starting at 494 nt upstream of the AUG, two would present in the 343 transcript and only the GGUUUGA motif would be present in the transcript starting at 61 nt upstream of the AUG. The HD+ mRNA appears to commence 494 nucleotide upstream of the AUG, thus the HD+ mRNA contains two regulatory motifs which are absent from the 61 mRNA. This implies that HD+ mRNA may be controlled by regulatory mechanisms which would not regulate the 61 mRNA. The inclusion of binding sites for splicing proteins in the 5'UTR of the longer more extensively spliced RNA suggests a role for the additional sequence in 494 transcripts.

Transcription termination can be effected by U-rich regions in the mRNA but these lie normally in the 3'UTR. Coupling promoter choice to site of transcription termination could in theory be achieved by a protein which recognizes sequences specific to a particular promoter and can associate with RNA polymerase II. Upon transcription (regulated by a different system?) this protein would bind to motifs in the nascent RNA and quickly associate with the polymerase. Termination could be achieved when a specific termination site associated with that promoter was recognized. Sequences that are present in both promoter and termination site might be a suitable shared signals. Intriguingly, two slightly degenerate copies of the 5'UTR SM binding sites motifs are present within 6 nucleotides of one another in intron 3, approximately 450nt beyond the 3'end of the hd-mRNA. These intron 3 motifs are almost identical to those in the 5'UTR; AgTTTTGTgG compared to AATTTTGTAG in the 5'UTR and tATTTTGA compared to

GATTTTGA in the 5'UTR (Figure 3.3B). The two 5'UTR SM binding motifs are present only in the 494 transcript which is produced by the promoter used for the HD+ mRNA. Similarly, 343 transcripts, which produce an HD+ transcript, contain one of these motifs. Production of HD+ mRNA requires read-through of the termination/polyadenylation signal within exon 3. The exclusive association of these two 5'UTR motifs with the HD+ transcripts and the location, within exon 3, in a region close to the end of the hd- transcript, of very similar motifs, links the 5'UTR to the site of transcript termination. The role of the exon 3 motifs could be determined by mutation, removal, substitution or alteration of spacing, in each case one could then determine the effect of these changes upon the quantity of HD+ transcript produced.

Formation of stem-loop structure could control the HD+ or hd- formation. Once transcription initiates at position -61 and while elongation occurs, a terminator protein could bind the DNA and follow the RNA polymerase causing termination in the intron 3 thus producing the hd- transcript. Alternatively if transcription initiates at position -494 stem-loop structure would form in this region blocking the site for terminator protein to bind. This would allow the RNA polymerase to continue transcribing till the end to produce the HD+ transcript. Interestingly, the human c-myc gene has two different promoters and transcripts originating from these have different termination sites [216]. Putative stem-loop structure formations in the promoter region were determined using CONTRAfold software (Figure 3.3C) [221].

(A)

>-494 to 0

GCTTTGCTGTTTTCTGGGTCGGAGACGCGGGTTCGGGGTCACTGCAAGACAAGATGCTGC
AAGTGGTGGTGGTGTGTGGGTTTTGTTGTACTGGGCCAGCATGATTTTGTTCGGAGGAT
TGC ACTGCAGCTGAACATTAAGACAGGCCACAT **GATTTGAAA**AGGAAAGGAAAAGACA
TAACGGCATCTCTGTTGGGATCCAGAATTGAGAAAGGGACGCTGGTTCGGTTTCCACAAT
GAATGAATGAAGGCCACCAGTGCATATGAGACTTTCTCTTCTATTTGTAGAGGATAAGT
CGGGGATTGAAATATATGTTCC**CCCAACAGAC**TGACAAGAGGAACTTAGAGCCTATGGA
CTGCCGATGAAATGGGAGAAGTACAAGGGTCTCGAA**AATTTTGTAG****TCGCCTGGCC**AAA
GAGCTGATGTCTAGG**ATTTTTTTCACCCAA**ATGCTGGCTGAATGCT**GTGAAATAAATT**A
AA**GTTTTGA**AGAAAGTGTTC**ATG**AAATATGGAGCGTCTGAATACAGCAGCTGCACAAG
CCTCGTCTTCGC

(B)

TCAATG**AGTTTTGTGC**CATAGTT**ATTTTGA**ACCATTAA

(C)

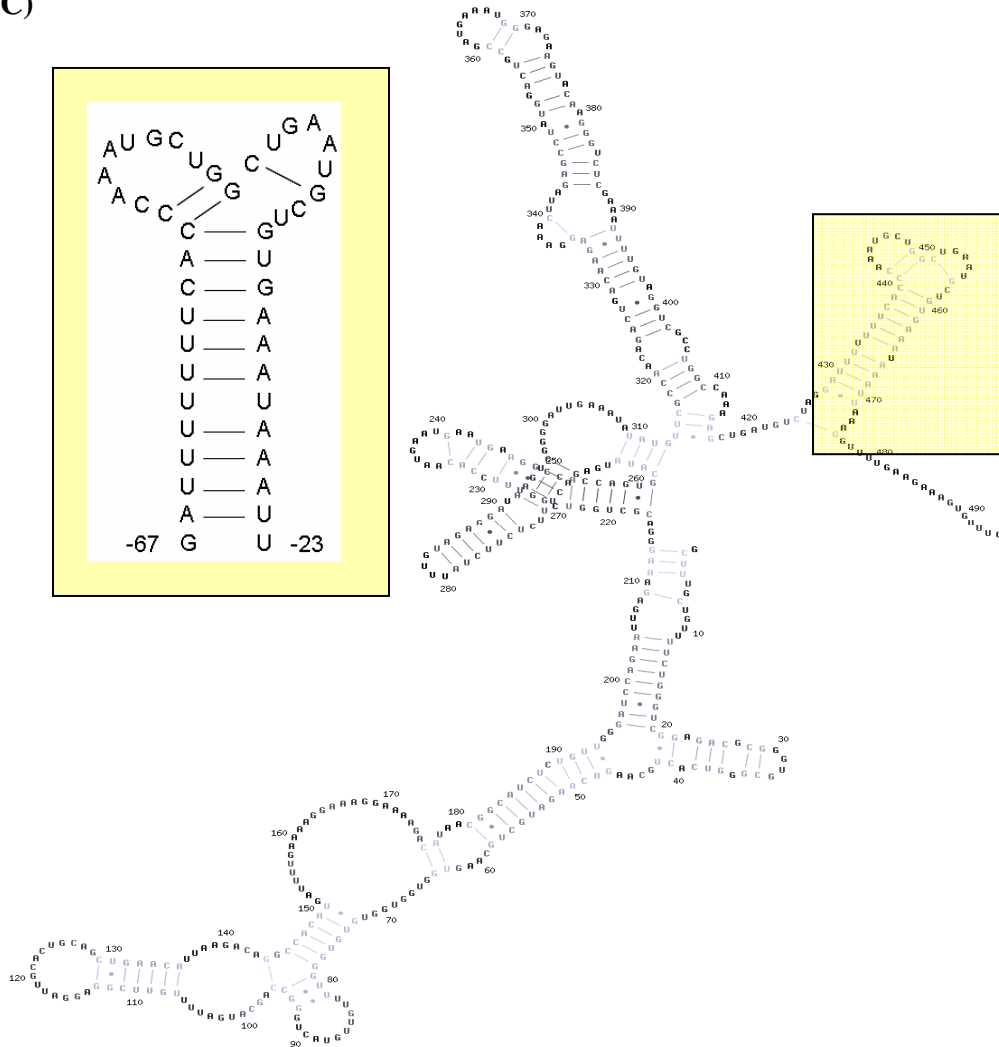


Figure 3.3 Transcription start sites in the promoter region and protein binding sites and stem-loop structures possibly involved in regulation.

(A) PtKN1 Transcription Start Sites. The three TSS used in the PtKN1 promoter are situated 494, 343 and 61 nucleotides upstream of the translation start site. The first 10 nucleotides of each of the transcripts are in boldface and underlined. Binding sites for three SM sites/small ribonucleoprotein (snRNP) are shown in red, blue and green. The 343 TSS corresponds exactly to the 5' end of the second SM binding site, thus the TSS of the transcript begins GATT.

(B) Protein binding site in intron 3. The sequence shown is 440 to 447 nt end of the hd-mRNA in intron 3.

(C) Stem-loop formation of promoter sequence. The stems of the putative 'stem-loop' structure highlighted in yellow box in the figure are shown in yellow with arrows below the sequence showing their orientation. The first 10 nucleotides of the stem loop structure 84 to 94 nt upstream of the ATG are highlighted in turquoise.

3.3.3 KNAT1-like Promoter analysis by phylogenetic footprinting

Phylogenetic footprinting involves comparison of conserved regions among orthologous regulatory sequence that are thought to be of functional importance [203-206]. Identifying the regulatory motifs by comparing the promoter sequence of KNAT1-like genes from different species will help in identification of conserved expression pattern [204]. Within conifers, paralogs of KNAT1-like genes have been identified as KN2, KN3 and KN4 and their divergence is suggested to have occurred before the split between *Pinus* and *Picea*. Promoter sequence of PgKN2 (*Picea glauca* KN2 - DQ259458) was found from NCBI database by blasting the PtKN1 promoter sequence. It has been suggested that KN1 and KN2 share more or less complete functional redundancy based on their position on the same linkage group being close and significant directional selection between them [96]. Hence, comparison of the promoter sequence of PtKN1 and PgKN2 will be helpful in determining regions of regulatory importance. In addition, various other promoters, such as *Arabidopsis thaliana* KNAT1 promoter sequence (AJ131822) [60], *Cardamine hirsuta* BP promoter sequence (DQ526379)

[207] which are dicotyledonous angiosperms *Zea mays* KN1 promoter sequence (AY312169) [208] *Triticum aestivum* wknox1d promoter sequence (AB182945) [209], which are monocotyledonous angiosperms, were used in the analysis. Promoters were analyzed using FOOTPRINTER [210]; <http://genome.cs.mcgill.ca/cgi-bin/FootPrinter3.0/FootPrinterInput2.pl>). This program takes multiple sequences as input and looks for regions of functional conservation amongst them based on phylogenetic relationship. I chose the word size of ten for motif search in order to avoid too many false positives when searching for smaller words. Two mutations were allowed in the 10bp motif. Programs such as FOOTPRINTER identify 50 or more motifs in any given sequence. To refine the search, I chose motifs that have been functionally characterized in other plants and/or are responsive to signals that are known to regulate knox genes. Several conserved regions, the regions of functional importance, were identified in these promoters (Figure 3.4). The sequence of these motifs searched against a database PLACE known regulatory motifs using a motif comparison program called STAMP [211, 212] (<http://www.benoslab.pitt.edu/stamp/>). The similarity of these motif sequences found by footprinter database to the cis-regulatory motifs found from PLACE database, their sequence and function are described in table 3.1. Cis-regulatory elements such as ethylene response element, element for endosperm expression, sugar response element, auxin response element, light response element were found in all KNAT1-like promoters. Cis-regulatory elements such as those found in anaerobic genes, element for root expression and poly a signal were found in pine, picea, *Arabidopsis* and cardamine KNAT1-like promoters and cis-regulatory elements such as WRKY transcription factor binding site, ABA response element, organ specific expression element, HD-ZIP

transcription factor binding site, pathogenesis related expression elements were found in maize and wheat KNAT1-like promoters. Many of these motifs were also found when searching for cis-regulatory elements in PtKN1 promoter. In general, promoters of maize and wheat have same motifs and promoters of pine, picea, *Arabidopsis* and cardamine have same motifs suggesting more promoter divergence between monocotyledonous and dicotyledonous angiosperms.

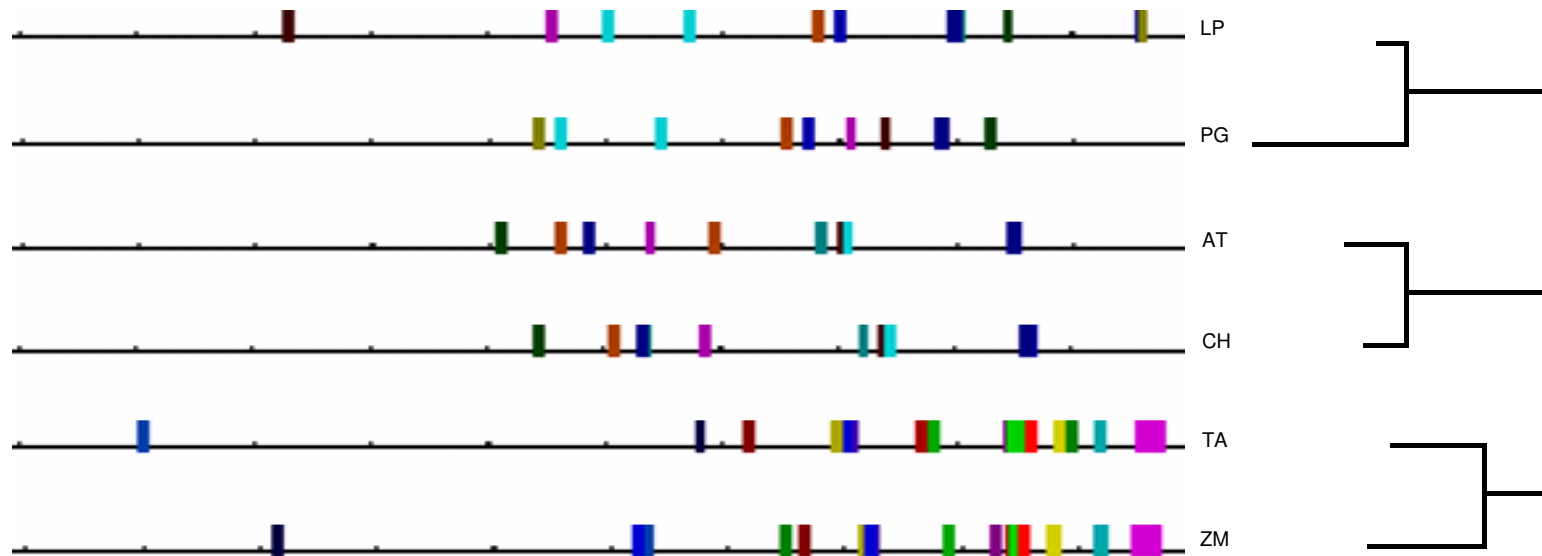


Figure 3.4: Promoter analysis by phylogenetic footprinting using footprinter program.

Phylogenetic footprinting involves looking for regions of functional conservation among related sequences. Promoter sequences from KNAT1 like genes were obtained from different species. Search was performed with word size 10, allowing 2 mutations.

The motifs are discussed with the same color coding in the next table.

LP – Loblolly pine, PG – *Picea glauca*, AT – *Arabidopsis thaliana*, CH – *Cardamine hirsuita*, TA – *Triticum aestivum*, ZM – *Zea mays*.

Table 3.1: Conserved regions found by footprinter in KNAT1-like promoters could be potential regulatory motifs.

The motifs found using footprinter program were searched against PLACE database for known regulatory motifs using STAMP motif comparison. The motif name and their function are described in the table.

Footprinter motif	PLACE motif	PLACE TFBS	E value	Function
ATTGAAATAW	ATTGAAATAW	ERELEE4	2.23E-07	ethylene responsive element
CTCTCTCCAA	TTGGAGAGAG	CTRMCMAMV35S	9.86E-08	Can enhance gene expression
AGCCTATGRR	AGCCTATGRR	Minus284MOTIFZMSBE1	4.04E-04	responsive for sugar responsive expression of sb1
CCTTCTCCTT	AAGGAGAAGG	TL1ATSAR	1.48E-05	consensus sequence overrepresented in pathogenesis related gene
TGCGCAGGAG	TGCGCAGGAG	INTRONLOWER	6.73E-05	consensus sequence for plant intron
CWTCTCTSTT	AASAGAGAWG	ANAERO5CONSENSUS	4.93E-07	motif found in anaerobic genes
ACTGTCTCTC	GAGAGACAGT	ARFAT	2.51E-08	auxin response factor
TSAAATATWT	TCAAATATWT	LECPLEACS2	4.68E-07	core element in Lecp binding cis-element
AAGAACTTCA	TGAAGTTCTT	YREGIONNTPRB1B	3.37E-09	required for ethylene induction
CTCTAATGAT	ATCATTAGAG	HDZIP2ATATHB2	1.94E-06	binding site of <i>Arabidopsis</i> homeobox gene
TTTATYTGRT	AYCAAATAAA	POLASIG1	4.23E-08	poly A signal
CTCCTTCCCC	GGGGAAGGAG	AMMORESIIUDCRNIA1	1.54E-04	nia1 transcription repression (chlamydomonas)
TKTGGKTTTT	AAAAMCCAMA	TELOBOXATEEF1AA1	3.62E-07	required for expression in root primordial
AAGTTCAAAG	AAGTTCAAAG	BOX1PVCHS15	2.43E-07	involved in organ specific expression in plant development

Footprinter motif	PLACE motif	PLACE TFBS	E value	Function
TGAAATAARA	TGAAATAAGW	TATABOX5	2.51E-08	TATA box
TTTGSTKTTT	AAAAASCAAA	AGTACSAO	1.23E-07	AGTA repeat binding protein binding site – silencing
CAAACCCTTT	AAAGGGTTTG	ELRE1PCPAL1	7.10E-10	light responsive element
TCTAATGATA	TCTAATGATA	GLUTEBOX1OSGT2	1.29E-06	for endosperm expression
GAGCTCGATC	GAGCTCGATC	NONAMERATH4	6.71E-05	nonamer motif
CCAAGAAGT	CCAAGAAGT	ABREMOTIFIOSRAB16B	3.44E-05	required for ABRE responsiveness
TCTCCTTCCC	GGGAAGGAGA	L1DCPAL1	5.96E-07	uv-b and protoplastization responsive element
CACCTCAAGC	GCTTGAGGTG	SBOXATRBCS	1.13E-07	sugar and ABA responsive element
GATTGACCT	AGGTCAAATC	WBBOXPCWRKY1	1.11E-08	WRKY protein binding site

3.3.4 In silico PtKN1 promoter analysis - Several cis-regulatory elements are found in KNAT1-like promoters

Putative cis-regulatory elements that could play a role in regulation of the PtKN1 gene were identified by bioinformatics approaches using Plant-CARE [213] (<http://sphinx.rug.ac.be:8080/PlantCARE/index.html>) and PLACE [214] (<http://www.dna.affrc.go.jp/htdocs/PLACE>) databases. The cis-regulatory elements found in PtKN1 promoter sequence and their functions are described in table 3.2 and table 3.3 for PLACE and PLANTCARE database respectively. Several cis-regulatory elements involved in light response, response to several hormones, several transcription factor binding sites, organ specific expression elements etc. were found in PtKN1 the promoter. Since a number of false positive motifs are generated in the analysis, phylogenetic footprinting was performed to eliminate them. I chose a few to compare them to the other KNAT1-like promoters mentioned above. Individual KNAT1-like promoter sequences were searched for the presence of these motifs; the presence or absence of these motifs in other KNAT1-like promoters, their position and number of occurrence was determined in an attempt to understand PtKN1 regulation and function. I made bar diagrams with those regulatory elements marked at their position relative to the length of promoter (Figure 3.5). The sequence and function of the motifs and the color codes used are described in table 3.5.

Table 3.2: Regulatory motifs found in PtKN1 promoter using PLACE database.

Several cis acting regulatory motifs and transcription factor binding sites were found in PtKN1 promoter when the sequence was searched against PLACE database. Their sequence and function are described in the table.

Motif name	Motif sequence	Function
Place		
CGCGBOXAT	VCGCGB	recognized by signal-responsive genes
ASF1MOTIFCAMV	TGACG	biotic/abiotic stress/auxin/Salicylic acid/Light induced regulation
WRKY71OS	TGAC	transcriptional repression of GA signaling pathway
SV40CORENHAN	GTGGWWHG	SV40 core enhancer
PRECONSCRHSP70A	SCGAYNR(N)15D	plastid response element in hsp promoters
ABRELATERD1	ACGTG	ABRE sequence required for induced expression of erd1
ACGTOSGLUB1	GTACGTG	endosperm specific expression
ACGTATERD1	ACGT	Etiolation-induced expression of erd1
CURECORECR	GTAC	copper response element
EBOXBNNAPA	CANNTG	bHLH factor binding site
MYCCONSENSUSAT	CANNTG	myc TF binding site in response to dehydration
CACGTGMOTIF	CACGTG	light induced regulation (G box)
RHERPATEXPA7	KCACGW	root hair-specific regulatory element
GTGANTG10	GTGA	pollen specific expression
AMYBOX1	TAACARA	GA responsive element
GARE1OSREP1	TAACAGA	GA responsive element
ARR1AT	NGATT	binding site for cytokinin-regulated TF
POLASIG1	AATAAA	polyadenylation signal
CACTFTPPCA1	YACT	mesophyll-specific gene expression

Motif name	Motif sequence	Function
DOFCOREZM	AAAG	binding site for Dof proteins
SORLIP1AT	GCCAC	light induced regulation
TATABOX5	TTATTT	TATA box
GT1CONSENSUS	GRWAAW	light induced regulation
EECCRCAH1	GANTTNC	(enhancer element) CO2 responsive, myb TF binding site
CIACADIANLELHC	CAANNNNATC	circadian control of gene expression
TBOXATGAPB	ACTTTG	light induced regulation
WBOXNTERF3	TGACY	transcriptional repression in response to wounding
SITEIIATCYTC	TGGGCY	meristem and anther specific expression
SORLIP2AT	GGGCC	light induced regulation
CAATBOX1	CAAT	CAAT box
MYCATRD22	CACATG	myc TF binding site in response to dehydration
MYCATERD1	CATGTG	myc TF binding site in necessary for erd1 expression
PYRIMIDINEBOXOSRAMY1A	CCTTTT	pyrimidine box - GA induced response
MYBCOREATCYCB1	AACGG	myb TF binding site
RAV1AAT	CAACA	RAV1 TF binding site
INRNTPSADB	YTCANTYY	initiator element
CATATGGMSAUR	CATATG	auxin response element
NODCON2GM	CTCTT	nodulin consensus sequence
OSE2ROOTNODULE	CTCTT	organ specific expression
ROOTMOTIFTAPOX1	ATATT	root specific expression
GT1GMSCAM4	GAAAAA	pathogen and salt induced gene expression
ANAERO2CONSENSUS	AGCAGC	motif found in anerobic gene

Table 3.3: Regulatory motifs found in PtKN1 promoter using PLANTCARE database.

Several cis acting regulatory motifs and transcription factor binding sites were found in PtKN1 promoter when the sequence was searched against PLANTCARE database. Their sequence and function are described in the table.

Motif name	Motif sequence	Function
PlantCARE		
CCGTCC box	CCGTCC	cis acting regulatory element related to meristem specific activation
ABRE	TACGTG, CACGTG	Abscisic acid response element
ARE	TGGTTT	cis acting regulatory element essential for anaerobic induction
CGTCA motif	CGTCA	MeJA responsiveness
G box	CACGTA, TACGTG	cis acting regulatory element involved in light responsiveness
GC motif	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GCN4 motif	TGTGTCA	cis-regulatory element involved in endosperm expression
I box	GATATGG	part of a light responsive element
Skn1 motif	GTCAT	cis-acting regulatory element required for endosperm expression
Unnamed 1, 3	CGTGG	not known
Unnamed 4	CTCC	not known
AAGAA motif	GAAAGAA	not known
GATA motif	AAGAATAAGG	part of a light responsive element
GC motif	CCCCCG	enhancer-like element involved in anoxic specific inducibility
TCA element	AAGAAGAATAA	cis-acting element involved in salicylic acid responsiveness
Unnamed 2	CCCCGG	not known

Motif name	Motif sequence	Function
BoxI	TTTCAA	light responsiveness
ERE	ATTTCAA	ethylene responsive element
GARE motif	TCTGTTG	GA responsive element
MBS	TAACTG	MYB binding site involved in drought inducibility
O2-site	GATGAAATGG	cis-acting regulatory element involved in zein metabolism regulation
TC-rich repeats	ATTTTTTCA	cis-acting element involved in defense and stress responsiveness
GT1 motif	GGTTAA	light responsive element



Figure 3.5: Regulatory motif search in KNAT1-like promoters from different species.

KNAT1-like promoter sequences from different species as mentioned in text were used in the analysis and of regulatory motifs were searched for using PLACE and PLANTCARE databases. The presence and position and number of occurrence of a selected number of motifs described in table 3.3 were determined in the promoters.

	Motif	Function
	BIHDOS1	BELL HD transcription factor binding site
	ABRE	Abscissic acid response element
	GARE	GA responsive element
	ARE	Cis-acting regulatory element essential for anaerobic expression
	Pyrimidine box	GA induced response
	GAG	Part of light responsive element
	SORLIP1AT	Light induced regulation
	ASF1MOTIFCAMV	Biotic / Abiotic stress / Auxin/Salicylic acid / Light induced regulation
	ATG	Start codon

Table 3.4: Regulatory motifs found in KNAT1-like promoters.

The colored boxes in Figure 3.4 are described in this table.

3.3.4.1 BIHD1OS site (TGTC A)

OsBIHD1 is a rice BELL HD transcription factor and is been shown to bind TGTC A sequence which is the characteristic cis-element DNA sequence of the HD transcription factor [222]. It is known that dimerization of BELL and KNOX is required for their function. It is possible that expression of some KNOX genes may be regulated by BELL transcription factors; and then upon activation when KNOX proteins are formed they dimerize with the same or different BELL proteins to regulate the expression of target gene. Two to four TGTC A binding sites were found in all the KNAT1-like promoters used in this study, except for in wheat wknox1. It is possible that the element is present in further upstream in the wknox1 promoter or it has some change in the binding site which is not recognized in the analysis.

3.3.4.2 ABRE (Abscisic acid response element; TACGTG, CACGTG)

ABRE is a cis-acting elements function in ABA-responsive gene expression under water-stress and seed desiccation conditions. ABA activates bZIP proteins to bind to ABRE sites and initiate transcription of ABA-inducible genes [223]. Under the influence of ABA, MYC and MYB transcription factors which are bZIP family proteins, activate transcription of dehydration responsive gene [224]. ABA levels in plants are also shown to be regulated by knox genes [153]; and in addition, AS1 (asymmetric leaves1) is a MYB encoding protein and has been shown to down regulate the expression of class I knox genes [225]. This suggests a possibility of cross-talk between ABA, KNOX1 genes and MYB transcription factor in order to regulate developmental changes. Thus the presence of ABRE motif in 5' flanking sequence of PtKN1 gene is consistent with these

results. ABRE motif was found only in pine, picea and maize KNAT1-like promoters but not the others.

3.3.4.3 GARE (gibberellin responsive element; TAACAGA, TAACAAA, CAACAGA, TCTGTTG)

Antagonistic relationship between GA and KNOX genes has been observed in previous research. KNOX genes STM and KNAT1 have been shown to down-regulate the expression of a GA synthesis gene GA 20-oxidase [159]. When KNOX activity increases GA signaling is reduced, and when KNOX activity decreases GA signaling is increased. This relationship between GA and KNOX gene has been observed in *Arabidopsis* and tomato [207], rice [153] and tobacco [156, 157] and is important for meristem function. The presence of GARE motif in 5' flanking sequence of the KNAT1-like promoters is consistent with the relationship between GA and KNAT1-like.

3.3.4.4 ASF1MOTIFCAMV (TGACG)

In response to auxin, salicylic acid or light promoter of genes containing TGACG elements are known to activate transcription of several genes. The presence of regulatory element that responds to auxin suggests a relationship between auxin and KNOX gene expression. Antagonistic relationship between auxin and KNOX genes has been observed when auxin transport was inhibited in tomato shoot apex and it caused expression of a KNOX gene LeT6 to highly increase [152]. Overexpression of rice *osh1* gene in tobacco causes a decrease in levels of auxin [153]. STM mutants show visible phenotypes due to ectopic auxin accumulation [57, 58]. Thus the presence of an auxin responsive motif in 5' flanking sequence of all the KNAT1-like promoters in this study indicates that a similar relationship between auxin and PtKN1 may exist.

3.3.4.4 Light responsive regulatory elements

Various light response elements are found in the KNAT1-like promoters. I have included SORLIP1AT (GCCAC) (Sequence over-represented in light-induced promoters) [226] which is present in all KNAT1-like promoters in this study and GAG motif (CATCTC) [227] which is present in pine, spruce, wheat and maize KNAT1-like promoters but absent from *Arabidopsis* and cardamine KNAT1-like promoters in our analysis. Expression of *knox* gene is affected by light; KNAT1 expression in light grown seedlings is less compared to dark grown seedlings. This is observed for KNAT2 as well, though to a lesser extent [60]. Many different types of light motifs are found in plant promoters. Since the GAG motif is absent from *Arabidopsis* and cardamine KNAT1-like promoters it is possible that other light motifs may be used in these promoters instead of the GAG motif. For example AF1 light response elements is found in *Arabidopsis* and cardamine KNAT1-like promoters [228].

3.3.4.5 Pyrimidine box

The pyrimidine box acts co-operatively with the GA response element and forms a GARC (giberrellin-responsive complex) to give high level of GA regulated response [229]. In addition, pyrimidine box and GARE have also been shown to be partially involved in sugar repression [230]. GARE and pyrimidine box are located close to each other in GA-responsive promoter [231]. Multiple GARE and pyrimidine boxes are found in the KNAT1-like promoters in this study, and the positions of one of each of these regulatory motifs to the other is close suggesting those motifs are most likely to be involved in GA related response.

3.3.4.5 ARE (Cis-regulatory element essential for anaerobic expression)

ARE was found in promoters of all KNAT1-like genes in this study, but KNAT1-like gene has not been shown to be involved in anaerobic response. The only *knox* gene shown to be involved in anaerobic response is KNAT4; its expression was found to be reduced under anaerobic conditions [232]. Only one ARE was found in each promoter except for cardamine which has 2 AREs but are at a distance of 94 bp. Experiments with maize *Adh1* gene which responds to anaerobic conditions have shown that 2 AREs are required in proximity to give anaerobic response and their orientation is also important [233, 234]. Therefore the presence of ARE element in KNAT1-like promoter may not be significant to its regulation or function.

Finding the same cis-regulatory elements in the KNAT1-like promoters from gymnosperm and angiosperm genes and from different databases gives confidence to believe that these cis-regulatory elements are present in the PtKN1 promoters and are involved in PtKN1 regulation. This study provides the base to conduct experiments in future to confirm the presence and regulatory roles of these cis-regulatory elements in PtKN1 promoter and suggests that KNAT1-like promoters are evolutionarily conserved in gymnosperms and angiosperms.

CHAPTER 4

CONCLUSION

I have isolated Knotted1 transcription factor from Loblolly pine (PtKN1(HD+)). By sequence similarity PtKN1 belongs to a family of KNAT1-like genes which are class I KNOX genes, identified in several plants and known to play a significant role in SAM development and maintenance [34, 117]. Like the members of KNOX family, PtKN1(HD+) has conserved regions such as the KNOX1 domain which is involved in suppression of its target gene [98], the KNOX2 which is involved in dimerization and has transactivation activities [99], the ELK domain might be involved in protein-protein interaction [34, 98, 99, 106] or it may act as transcriptional repressor [98], and the HD which is a DNA binding domain [73]. In addition I also isolated a novel form PtKN1(hd-) that lacks important domains ELK and HD. I have shown that both forms can express protein. My results from RT-PCR and western blotting over different stages of pine embryo development show that the PtKN1(HD+) and PtKN1(hd-) mRNAs and proteins are developmentally regulated. The semi-quantitative PCR and Western blotting procedures were repeated several times using similar or different conditions and the similar data was obtained suggesting that the results presented are reliable. Real-time PCR can be performed to confirm the results of the semi-quantitative PCR and with added advantage of providing statistical data.

By using a combination of PCRs and genomic walks, I have isolated several overlapping clones spanning almost entire length of PtKN1 gene. In addition several overlapping genomic sequences in intron 3 region were found from NCBI database. 5`

and 3` end contigs were made using the isolated and database sequences which were 5751 and 1943 bp in length respectively (Figure A.5 in appendix A). A total of 7.7kb PtKN1 gene sequence was obtained of which 4.8 kb was intron 3. The intron 3 sequence is the largest part of knox genes; 5kb in maize (*Zea mays*), 5.5kb in rice (*Oryza sativa*) and over 4 kb for KNAT2 in *Arabidopsis* [60, 90, 193] suggesting that the intron 3 sequence of PtKN1 gene is almost complete. A small portion of intron 3 remains to be cloned.

The PtKN1 gene sequence that has been obtained suggests that PtKN1(HD+) and PtKN1(hd-) mRNAs have been produced by alternative processing of a single gene. Like the genes of KNOX family, PtKN1 gene has 5 exons and 4 introns the exon 1, intron 1 and exon 2 regions are common for both forms. The beginning of intron 3 is the region that codes for the (hd-) specific mRNA sequence i.e. the 3` ORF region and the 3` UTR. The exon 4 and exon 5 regions code for the (HD+) specific mRNA sequence i.e. the ELK and HD regions and the 3`UTR. Even though the available sequence suggests that there is a single PtKN1 gene, the possibility that PtKN1(HD+) and PtKN1(hd-) are separate genes of a gene family cannot be ruled out because of the gap in intron 3 region which remains to be cloned. However, complete cloning of a contiguous genomic sequence which encodes both HD+ and hd- sequences in a single gene would not rule out the possibility of a second hd- gene somewhere in the very large and complex loblolly pine genome. But the following experimental findings lead to the same suggestion that PtKN1 gene has a single structure in the loblolly pine genome.

- 1) Twenty nine independently isolated PtKN1 cDNAs were nearly identical in sequence, with a minimum of zero to a maximum of eight nucleotide differences over

their length. Ten out of twenty nine PtKN1(HD+) sequences have G instead of C at position 118 (Supplementary Data Figure A.14); T instead of A at position 310 (Supplementary Data Figure A.15) and thirteen PtKN1(HD+) sequences have T instead of C at position 579 downstream of the start codon (Supplementary Data Figure A.16) suggesting that these are alleles. Five PtKN1(hd-) cDNAs were nearly identical to one another and to the overlapping regions of the PtKN1(HD+) cDNAs. Nucleotide differences are few and random, suggesting that these may be sequencing errors.

2) Eleven PtKN1(hd-) genomic clones covering exon 1 through the proximal end of intron 3 showed 99.6% identity (1 to 9 nucleotide random differences over 2090 bp). The exonic regions of these clones show 100% nucleotide identity with (HD+) and 100% nucleotide identity with (hd-) cDNAs.

3) Restriction enzyme digestion patterns for fourteen independent genomic clones of the region from the 5'UTR to exon 2 were identical; similarly fourteen independent clones of the region covering exon 4, intron 4 and exon 5 show identical digestion patterns (Figure 2.6). Clones 8 and 14 in Figure 2.6B and Figure 2.6C and clone 10 in Figure 2.8E and 2.8F show a different pattern of cutting but they are consistent with the same gene structure, the insert being in opposite orientation in the vector.

4) Sixteen independent genomic clones containing approximately 945 bp of the PtKN1 promoter were generated from a region within the open reading frame that is common to both types of transcript. The sequences were identical, containing between zero and eight random nucleotide differences over their length. Eight additional independent genomic clones which extended the region to 1670bp beyond the TSS were identical in sequence.

As described above, the identity among the cDNA and genomic sequences of PtKN1(HD+) and PtKN1(hd-) is very high when compared to what the identity among genes seen in many gene families in plants. This identity extends to the intronic regions which are also identical among genomic clones. Overall, these data argue for the PtKN1 gene is present in a single copy, and that both types of transcripts are produced by this one gene, rather than being produced by several distinct genes whose structure reflect. The sequence similarity among PtKN1 clones is far greater than is observed among different members of gene families. For example the Adh gene family in *Pinus banksiana* has seven genes which show identity of 75.4% to 98.7% in the exonic regions. The sizes of their introns differ due to direct or inverted repeat regions and further their 3'UTRs show a huge sequence divergence; too great to permit construction of any meaningful alignment [235]. The HSP100 gene family in *Arabidopsis thaliana* has seven genes which show identity of 21% to 70% [236]; and in rice has five genes which show identity of 32% to 48% [237] in exonic regions. The examples given above indicate that the nucleotide sequence similarities among different members of gene families are usually confined to the exonic regions or to particular domains. Gene regions which do not encode functional domains vary considerably. Strong sequence conservation is more usually observed at the amino acid level. Exons and regions corresponding to untranslated regions of the mRNA, are under much less selective pressure. When homologous genes are compared, greater differences in nucleotide sequence are seen in the comparable introns, than are seen in the comparable exon regions. In the light of these data, the sequence identity along the entire length of cDNA and genomic clones is most plausibly explained by a model where PtKN1 is a single gene which produces both

PtKN1(HD+) and PtKN1(hd-) mRNA forms by alternative processing of the primary transcript.

The structure of the gene indicates that the PtKN1(hd-) mRNA could be produced by splicing of the first two introns followed by cleavage-polyadenylation or termination of the primary transcript within intron 3 [196, 238, 239]. If a strong termination signal is present in intron 3, then the transcription machinery would never be able to go beyond the termination point to produce the pre-mRNA including the HD+ specific coding sequences and hence PtKN1(HD+) transcript would never be produced. Therefore presence of a strong termination signal is ruled out. Alternatively if a weak termination signal is present at the end of PtKN1(hd-) transcript coding region the in intron 3, then the hd- transcript could be produced by pause dependent termination in intron 3 [240]. Motif search against the UTRscan database [241] (<http://www.ba.itb.cnr.it/UTR/>), in the intron 3 region did not show presence of any polyadenylation or termination signal. But this does not rule out the possibility of a polyadenylation or termination use to produce the hd- transcript since the databases of regulatory motifs may not be complete due to the lack of complete genome sequencing for many eukaryotic species.

I identified the two potential transcriptional start sites in the PtKN1 gene by using 5'-RACE to map the 5'-end of mRNAs. By using a transcript-specific primer I distinguished the transcriptional start sites of the HD+ and hd- transcripts and demonstrated that all HD+ mRNAs were transcribed for the 5'promoter, producing an mRNA which began 494 nt upstream of the ATG, while all the hd- mRNAs were transcribed from a point 61nt upstream of the ATG. Regulatory mechanism involving transcription factors could control the formation of HD+ and hd-. Activation of a certain

transcription factor or protein and their binding in the region promoter and intron 3 region could result in formation of hd- transcript. Binding in the -494 to -61 region would block the polymerase binding site for HD+ transcript. Binding in the intron 3 region would block the RNA polymerase machinery to move further and terminate the transcript hence producing the hd- form. Alternatively if the transcription factor binding sites are not occupied the RNA polymerase would bind the initiation site for HD+ and continue to transcribe till the end to produce the HD+ transcript.

Most interestingly I report the presence of three sequences within the 5'UTR to which have the consensus sequence of binding sites for SM protein/snRNP proteins involved in RNA processing. All three motifs are present in a transcript starting 494 nt upstream of the AUG but the transcript starting 61 nt upstream has only one motif. The distribution of the motifs may have some regulatory significance for the two mRNAs. Subsequently I identified two motifs which are almost identical to the motifs found only in the 5'UTR of the 494 transcript, located only 6 nucleotides apart in intron3. The intron 3 motifs are located only 450bp downstream of the end of the hd- mRNA. The exclusive association of two 5'UTR binding motifs for SM proteins with the HD+ transcript and the location, within exon 3, in a region close to the end of the hd- transcript, of very similar motifs, links the 5'UTR to the site of transcript termination. A regulatory mechanism could thus be envisioned whereby proteins which recognize particular 5'UTR motifs will also be able to recognize essentially the same motifs in intron 3 at a point where transcription termination or cleavage-polyadenylation occurs. Binding of these proteins at this point would in some fashion prevent transcription termination and the RNA polymerase would continue transcribing the gene and would produce a transcript

which would ultimately be processed to produce HD+ mRNA. This evidence for this mechanism is at best circumstantial; however it does, for the first time, present a plausible mechanism by which transcripts initiated from a particular promoter could identify a particular terminations site. The hypothesis is readily accessible to experimental verification and certainly seems worth examining.

I have isolated a spliced form of KNAT1 in *Arabidopsis* which lacks ELK and HD. This spliced form is only expressed in *Arabidopsis* seeds 5 day after fertilization and rosette leaf tissue, but is absent in *Arabidopsis* seeds 10 day after fertilization and seedling tissue. The expression of HD spliced KNAT1 mRNA is very low compared to the full length KNAT1 mRNA. Our results from Western blotting show that this form expresses protein and consistent with the mRNA expression I also see KNAT1(hd-) protein expression only in rosette leaf but not in seedling tissue. I have isolated spliced form of KNAT1 in *Arabidopsis* which lacks MEINOX encoding sequence. This spliced form is expressed in *Arabidopsis* seeds 5 and 10 day after fertilization and in seedling tissue and the level of expression is very low compared to the expression of the full length KNAT1. My results from Western blotting show that this form expresses protein and consistent with the mRNA expression I also see KNAT1(knox-) protein expression only in seedling but not in rosette leaf tissue. The fact that the expression pattern of KNAT1(hd-) and KNAT1(knox-) is different suggests different forms of KNAT1 regulation. By database search, I have identified spliced forms lacking HD and MEINOX regions in monocots such as maize and rice respectively. I have identified a splice form of maize gene rough sheath1 (rs1) mRNA which lacks ELK/HD encoding regions and a

splice form of rice OSH1 mRNA which removes some of the MEINOX encoding sequences for rice OSH1.

Based on what is known about the role of conserved regions of KNOX proteins, I suggest that the HD or MEINOX spliced forms may act as important regulators of knox gene function. Any protein-protein interaction possible through the ELK domain will be lost and without the HD the hd- form will not be able to recognize and bind DNA or localize to the nucleus. This implies that the hd- protein may function cytoplasmically, perhaps pairing with proteins thereby preventing their import into the nucleus. Alternatively, hd- protein could function in the nucleus by acting as a dominant negative regulator of the HD+ form by competing with its NLS containing KNOX or BELL interacting partner. KNAT1(knox-) may not be able dimerize with other protein; but they may be able to bind DNA with low affinity as a monomer [201], thus blocking the sites for KNOX homo or hetero dimers to bind and function.

I have isolated the promoter region of PtKN1 gene, 1730 bp upstream of the start codon. The TSS of PtKN1(hd-) is 61 bp upstream of the start codon. PtKN1(HD+) has 2 different TSSs which are 343 and 494 bp upstream of the start codon. The different TSS in the same promoter region suggests a possibility of cell or tissue specific gene regulation [215, 216]. Since the TSSs for PtKN1(HD+) and PtKN1(hd-) transcripts are different, it is likely that the choice of the TSS determines which form will be produced [216]. Interestingly, the human c-myc gene has two different promoters and transcripts originating from these have different termination sites [216]. I performed in silico analysis of the promoter sequence by employing phylogenetic foot printing. I looked for regions of functional conservation in orthologous KNAT1-like promoters and found cis-

regulatory elements which could possibly regulate the expression of PtKN1 gene. I found several cis-regulatory elements such as hormone response elements, transcription factor binding site, organ specific expression motifs which have been previously shown to be related to KNOX1 gene regulation. This suggests that the regulatory mechanisms that control KNOX genes expression in angiosperms are also present in gymnosperms.

CHAPTER 5

FUTURE DIRECTIONS

Having completed isolation and an initial molecular characterization of PtKN1 forms it is important to determine their functions. One approach would be to make transgenic *Arabidopsis* plants over-expressing PtKN1(HD+) and PtKN1(hd-) forms to determine their function. The ease of transformation and the availability of mutants make *Arabidopsis* an attractive host plant for a series of analyses. When HBK3 (Norway spruce Knox (HD+)) was expressed in *Arabidopsis* several phenotypes observed resembled those produced by overexpressing KNAT1, however a number of post embryonic phenotypes were observed such as lobed rosette leaves and dwarf phenotypes, characteristics that were not observed for spruce [71]. Markel et al. (2002) deleted different regions of the *Arabidopsis* KNOX protein, STM, and performed functional studies in transgenic plants. Overexpression of a KNOX protein with ELK and HD regions deleted would be similar to overexpressing a HD-lacking protein. When the MEINOX domains of either STM or KNAT1 were overexpressed in *Arabidopsis*, *stm* or *knat1 loss-of-function* phenotypes such as elongated internode and reduced leaf size were observed. These results were suggested to be due to squelching mechanism, which depletes the wild-type protein from its interacting partners [242]. Overexpression of a HD-lacking protein KNATM-B in *Arabidopsis* under the influence of an inducible promoter, causes late flowering and formation of elongated petiole and serration in leaves [113]. The experiments of Markel et al. (2002) and Magnani and Hake (2008) examined the effects of fragments of KNAT1 which in many respects resemble the hd- proteins that

we wish to investigate. These results indicate likely phenotypes arising from overexpression of hd- proteins.

If the constitutive overexpression of PtKN1(hd-) mRNA is fatal, no transformants would be obtained. In that case, hd- could be put under the control of an inducible promoter such as nopaline synthase which is inducible by auxin [243]. The roles of the hd- transcript at different times of development could be investigated by inducing expression at different times during development and observing the range of phenotypes that are generated. Following induction, tissue can be collected and changes in gene expression observed using microarray. The PtKN1 clone could be cloned downstream of the cauliflower mosaic virus 35S promoter and the construct be then introduced into the *Agrobacterium tumefaciens*. The resulting *Agrobacterium* strain can be transformed into *Arabidopsis* by floral dip method; by dipping the *Arabidopsis* plant in *Agrobacterium* that are suspended in a solution containing a sugar and a surfactant [244]. Transformants can be selected on kanamycin and the phenotypes of plants grown can be studied. Transformants can be obtained at a higher rate by the dipping method than by the original infiltration protocol used for *Agrobacterium* transformation which is labor-intensive [245]. In an attempt to understand role of PtKN1(hd-) during embryogenesis, it can be overexpressed in loblolly pine embryos. Overexpressing hd- in pine embryos may produce embryos which look normal or are slightly misshapen or very misshapen or fail to complete development or germinate poorly or produce stunted plants or a range of other things. The large sizes of loblolly pine embryo facilitate molecular studies in pine which are difficult in *Arabidopsis* embryos [28, 246]. Transformation of mature pine

zygotic embryos using *Agrobacterium* [247] and particle bombardment [248] have been successfully used and reported.

To show loss of function phenotypes of hd- form, antisense or siRNA constructs can be created and used against hd- mRNA. Antisense construct for PtKN1(hd-) can be made and placed downstream from a ubiquitin promoter. PtKN1(hd-) in sense orientation can be used as control. A region specific the PtKN1(hd-) can be selected and amplified in PCR using primers with restriction sites at the ends. The amplified products can then be ligated into a vector with same restriction ends and downstream from an ubiquitin promoter. A 21 nt region specific to PtKN1(hd-) and starting with AA should be selected to be the target sequence for siRNA construct. Oligonucleotides in sense and antisense orientation in hairpin structure can be ordered from ambion (Austin, TX). The hairpinRNA can be cloned into a pHELLSGATE vector, under the control of pOp6 promoter and synthetic transcription factor LhGR. This vector system is dexamethasone-inducible [249]. The vector construct can be transformed into pine using particle gene gun. The gold particles used for transformation can be coated with sense or antisense PtKN1(hd-) plasmid constructs. The transformants can be selected for a resistance gene under the control of ubiquitin promoter. Such constructs under control of ubiquitin promoters have been successfully transformed by particle bombardment into spruce embryos [250]. Once the transformants are selected, the embryonic growth and morphology can be examined. Antisense constructs have been made and expressed successfully in Norway spruce somatic embryos and morphological differences have been observed [71].

Two HD lacking KNOX isoforms have been reported in angiosperms; KNATM gene in *Arabidopsis* [113] and PTS gene in tomato [251]. Both are separate genes lacking HD sequences, and are not expressed from spliced mRNAs produced by a gene which contains HD sequences. The work described in my dissertation is the first report of KNAT1-like mRNA and protein lacking HD or MEINOX domains produced through alternative spliced products of the same gene. Future research to isolate and characterize such hd- and knox- forms from different plant species is important to understand its function and evolution. Comparison of sequences of HD lacking mRNAs and proteins, and the gene structures may provide information about when the form originated during evolution. Database search can be performed to find the sequence for hd- and knox- forms in other plants. Spliced forms of some species may have excise the exact same HD or KNOX sequence as seen for KNAT1(hd-), RS(hd-) and KNAT1(knox-) however variations where domains are partially removed are also possible. Spliced forms of some other species (gymnosperms) may have a small region of unique sequence that is not seen in the canonical form as in case of PtKN1(hd-) form. Because of such differences, any sequences obtained through database search will be of great help in actual isolation of the clone from that species. In case of spliced forms like KNAT1(hd-) which does not have any unique sequence, PCR with two gene-specific primers can be used to isolate the hd-cDNA. In case of spliced forms like PtKN1(hd-) which has a unique sequence at the 3` end, 3`RACE PCR will be a good way to isolate the hd- cDNA. Use of two gene-specific primers has an advantage of specificity over the use of one gene-specific primer and an adapter primer in RACE reaction; because of the unique region at 3` end, the former method is not helpful.

The 3' end of PtKN1(hd-) mRNA codes for 39 amino acid which are unique to the hd- protein. Whether or not the 39 amino acid region has any function can only be determined experimentally by expressing a PtKN1(hd-) protein that lacks this region. In order to express such protein, PtKN1(hd-) cDNA needs to be amplified in PCR in such a way that the C-terminal region coding for the unique 39 amino acids is not included. The primers for the PCR can be designed such that they have restriction sites at the ends and can be cloned directly in correct orientation in expression plasmid (such as pET-28 from Novagen, San Diego, CA). Once the plasmid is prepared it can be introduced into the *Agrobacterium tumefaciens* and transformed into *Arabidopsis* using same protocol as described earlier in this chapter. If the region including the 39 amino acids is a conserved and has function, phenotypes will be observed due to its overexpression. In future when more PtKN1(hd-)-like clones will be available from several plants; phylogenetic analysis by doing sequence comparison may provide some information about the conservation of this unique region and shed light on its function.

I have shown in this study that both the PtKN1 forms are translated. The next question that arises is to determine what proteins interact with the PtKN1(HD+) and PtKN1(hd-) forms. In-vitro pull down assay or yeast two hybrid assays can be used to determine the interacting proteins. Pull-down assays have been used successfully to determine interaction of rice protein OSH15 with other KNOX proteins [98]. Yeast two hybrid assay has been used successfully to show interaction of HD+ KNOX and BELL proteins [93, 100-102, 104, 105, 125] and interaction of HD-lacking KNOX protein KNATM-B and its isoforms with BELL proteins and other KNOX proteins [113]. PtKN1(hd-) protein can be used as bait and BELL proteins known to interact with KNOX

proteins as prey in yeast two hybrid assays; and the strength of interaction can be assessed using reporter gene such as lacZ. Using appropriate vector (such as pGEX from Amersham Biotech, Piscataway, NJ), fusion proteins between glutathione S-transferase (GST) and PtKN1(Hd+) or PtKN1(hd-) proteins can be generated. The GST-PtKN1(HD+) or GST-PtKN1(hd-) will serve as bait and total protein isolated from pine embryos containing interaction target proteins as prey. Glutathione agarose beads can then be used to capture the GST-tagged PtKN1(HD+) or GST-PtKN1(hd-) fusion protein with their interaction partner. The advantage of pull-down assay is that the target protein does not need to be known, and therefore new protein partners if any for the PtKN1(hd-) protein can be determined.

My RT-PCR analysis shows that both the PtKN1 forms are expressed at all stages of embryo development. It has been known from previous studies that knox genes are involved in meristem formation and maintenance of its function. The next question that arises therefore is whether PtKN1 shows same expression pattern. A closer look at expression in different tissues can be gained by dissecting embryos into different regions such as shoot, hypocotyl and root regions will help locate the exact regions of PtKN1(HD+) and PtKN1(hd-) expression. More specific tissue expression can be viewed by in situ hybridization using probes specific to hd- regions. Careful designing will be required to create HD+ and hd- specific probes. Since the 3' ORF and 3'UTR of hd-mRNA are unique, probe can be generated in that region and it will be hd- specific. The 3' ORF sequence of PtKN1(HD+) mRNA is not present in PtKN1(hd-) mRNA. But the ELK and HD regions cover most of the 3'ORF of PtKN1(HD+) mRNA which are conserved in KNOX family. Therefore a probe in 3'ORF of PtKN1(HD+) mRNA may

bind to all KNOX mRNAs depending of the exact region of probe binding. Therefore a probe in the 3' UTR of PtKN1(HD+) mRNA will be a better choice. Once the regions of probe binding are determined for PtKN1(HD+) and PtKN1(hd-) mRNAs, using two gene-specific primers those regions can be amplified in PCR. The amplified products can be put in cloned in vector downstream of T7 or SP6 promoter in antisense orientation and reverse transcribed to generate the probe. Sense probe should be used as control. The probe can be labeled with digoxigenin or flouorescein for detection. In situ hybridizations have been done successfully in our lab [189] and have also been done for two HD genes PaHB1 and PaHB2 in Norway spruce [191].

Isolation of exon 3 of the PtKN1 gene has to be completed in order to gain thorough understanding of the mechanism responsible for alternative splicing, that produces either PtKN1(HD+) or PtKN1(hd-). Further experiments would determine if PtKN1(hd-) is produced by cleavage-polyadenylation or termination within intron 3. The competition between polyadenylation/termination and splicing can be shown by partially inhibiting splicing in which case more PtKN1(hd-) will be produced; or by partially inhibiting cleavage-polyadenylation and termination in which case more PtKN1(HD+) will be produced. Antisense oligonucleotides against intron sequences [252] or splice sites [253] have been successfully used to inhibit splicing. It has been shown successfully in tobacco that insertion of U-rich elements within the intron increases splicing and replacement of one U with G reduces splicing by 50%, due to formation of hairpin structures in intron [254]. Similar mechanisms can be used to inhibit splicing of PtKN1 gene to increase the formation of hd- transcript. By mutating specific regions or creating

several deletion constructs within the promoter it could be shown if the choice of the TSS determines which form of PtKN1 will be produced.

After isolating the promoter region of PtKN1 and identifying putative cis-regulatory elements in this region the next important thing to determine is the regions of functional significance. The next experiment to be done would be to create several GUS constructs with different regions of PtKN1 promoters being deleted to determine the region that would play a role in regulation of the PtKN1 gene. Since the 1730bp PtKN1 promoter sequence was obtained from two different genomic walks, primers at the ends of their contig sequence can be used to generate a single molecule clone in PCR. If restriction sites are added at the end of the primers used in PCR, the products can be cloned into an appropriate promoterless vector (such as pBI101 from Clontech, Palo Alto) upstream of a GUS cassette. The promoter-GUS constructs can be introduced into *Agrobacterium tumefaciens* and then transformed into pine embryos. PtKN1 expression pattern can be determined by tracking GUS staining pattern. 5' and 3' deletions can be constructed by PCR amplification using primers in appropriate regions. Three different TSSs were determined, one for PtKN1(hd-) mRNA and two for PtKN1(HD+) mRNA. Therefore creating 3' deletions will be helpful in determining if any regulatory motifs involved in level or localization of expression are present in these regions. GUS constructs have been made and used successfully for PtNIP1 promoter which is expressed during pine embryogenesis [255]. Walking further upstream to get more sequence may be required to determine the elements present in the distal regions of promoter.

APPENDIX A

SUPPLEMENTARY TABLE AND FIGURES FOR CHAPTER 2

Table A.1: Primer used for genomic sequence cloning.

Forward Primer	Primer sequence (5' to 3')
KF	TTGTAGGTCGCCTGGCCAAAGAGCTGA
KI3U1	AGGGCCATGGACGATATCGAGGGAT
KI3U2	GGGCTTACAGGAAACCGCTGGGAGGTTAGT
KSF1	ACAATGAGAAGGAAGATTGCGCCTTACA
KSF2	TTGAAGGGATTCATTTATATTGGGCCATTA
KEF3	AGAACGGCAGTGGTGGTGAAACGGAT
K3U1	CGCCACTGGAAGCCTTCTGAAGATATGCAC
K3U2	TCACAGTCCTCACAGTGCTGCCTTGTATG
Reverse Primer	
KI5R1	ACGTACTACAACCATAAGTGCCACTCACTA
4R1	TCCTTGATTGTGGCGCCATGAGAAATGGATGCAA
4R2	GCATGATGATGTTTAGGGTGCTAACTTGGG
P5R1	ACTCCCTATCCTCTACTGCATGGTGATCCACTTCCTG
P5R2	TCACCACCACTGCCGTTCTCGACTTCCTCCGA
KR	GTTTTTCAGCCGAAGGTTCTTTCTAGCAAT

GAGAAGTACAAGGGTCTCGAAATTTTGTAGGTCGCCTGGCCAAAGAGCTGATG
TCTAGGATTTTTTTTACCCAAATGCTGGCTGAATGCTGTGAAATAAATTAAG
GTTTTGAAGAAAGTGTTTCATGAAT

132 **atg**gagcgtctgaatacagcagctgcacaagcctcgtct
M E R L N T A A A Q A S S
171 tcgctttatggagttaacatggcagaatatggagatgctggcgtc
S L Y G V N M A E Y G D A G V
216 agctcgatgatggggttgataaccaacatgagcctgagcaaagc
S S M M G L I T Q H E P E Q S
261 gaaaatattatgtccacgaggatcccaagctcgttttcatcattc
E N I M S T R I P S S F S S F
306 catggccatgccgattgccttctctcagcagcaatgtttcagggt
H G H A D C L L S A A M F Q G
351 tctcaaggagatcataagctcaatccacagcctgggatgaaccag
S Q G D H K L N P Q P G M N Q
396 cagctagtctctgagcagcttatcatgtcagattcgtccatgccg
Q L V S E Q S I M S D S S M P
441 ttgggttaagacaaaagcttgctctgggtcttcgtaatcagtttgaa
L V K T K A C S G L R N Q F E
486 tttcacaggaacaacccggaaattgctacacagatcagtcctca
F H R E Q P G N C Y T D Q S S
531 aatattccgctaagccccatagtcacaccgttagcctcgcaggct
N I P L S P I V T P L A S Q A
576 cgaggagaagcgcggatgataccgtccttggatgccaacggtgct
R G E A R M I P S L D A N G A
621 catttcaatgtggataacgaggagcatgcaataaaatcgaaaatc
H F N V D N E E H A I K S K I
666 ttagcgcacccacagtatccgagcttggtgggagcttacatcgac
L A H P Q Y P S L L G A Y I D
711 tgtcaaaagattggggctcctccagaagctgtggcccgtttgat
C Q K I G A P P E A V A R L D
756 gctctaactcgcgaacaccaagaccacacagtcggactgtgagc
A L T R E H Q D P Q R R T V S
801 atcggaatggaccagaactcgatcaatztatggaggcataactgc
I G M D P E L D Q F M E A Y C
846 gagatattgactaagtaccatgaagagcttgccaaacctttcaa
E I L T K Y H E E L A K P F K
891 gaagcaatgctgtttttgaagaagattgaaacccagttcaactcc
E A M L F L K K I E T Q F N S d
936 ctgggaaaaggaacgataagaatcttctccagctgacgatgac
L G K G T I R I S S P A D D D
981 gaaaagacggaggggaggtggatcttcggaggaagtgcgagaacggc
E K T E G G G S S E E V E N G
1026 agtgggtggtgaaacggattttcaggaagtggatcaccatgcagta
S G G E T D F Q E V D H H A V
1071 gaggatagggagttaaaaaatcatctccttcgcaaatacagtgga
E D R E L K N H L L R K Y S G
1116 tatttgagtagtctaaaacaggaattcatgaagaagaaaaagaaa
Y L S S L K Q E F M K K K K K
1161 ggaaagctccctaagatgcacggcaaaagttactcgattggtgg
G K L P K D A R Q K L L D W W
1206 agtctgcacgacaagtggccatataccttcggagacggagaaaata
S L H D K W P Y P S E T E K I
1251 gctttggctgaatgcacggggttgatcaaaaacaaataaataat
A L A E C T G L D Q K Q I N N
1296 tggtttataaaccaagaaaacgccactggaagccttctgaagat
W F I N Q R K R H W K P S E D


```

1341 atgcacttcatggtaatgaacagtcacagtcctcacagtgctgcc
      M H F M V M N S H S P H S A A
1386 ttgtatggtgagagacatatgatgactgaagggatctttag 1427
      L Y V E R H M M T E G Y L *
ATTGCTAGAAAGAACCTTCGGCTGAAAACAGCACACAATGCTATTGCTTTTGT
TGTATTTAATTGGCATGGCTTTCAATTTTCTTTCCCTTTTTTATTGGTAAGTA
AAAACCTCAAATTTATATAAAAAGGAAAGATAAAAAATAGTAATGACCCATAT
GCAAAGAAAAAGGTTTAAAAGAGTGGGGAGTTACACACAGCCATAAAATGCCA
GATATCGAGCAGGTACAATTCGTAAACTAAAAGGATAGGAGAAGAAAACAAAT
TTAAAAGGGAAATAAAGTATAACACTTCGGCACTAGATCAGGTAGTCCTAATT
GCCGGGAGCAGTAAACCTAACGGTCAAGCTAGCGTTAAACTAGTAAACCCCT
AACGGCCAAGCTAGCGTTAAACTGAACGAGCTTAACACAGAGGCATTGTGTCA
ATCGCTTCTGGGCCCTGGCACTCTCTCTCCCTCAGACGATATTGAAGAAGCTA
AGTTAAAGTTGTATGCACACTTCGTAATTTTCTGACTTTGAAGTAGTATCTCA

```

Figure A.1: Nucleotide and deduced amino acid sequence of PtKN1(HD+).
The beginning of the sequence is the 5` end of PtKN1 cDNA. The deduced amino acid sequence is shown below the nucleotide sequence. The start and stop codons are highlighted in bold. The end of the sequence is the end of the cDNA.

GAGAAGTACAAGGGTCTCGAAATTTTGTAGGTCGCCTGGCCAAAGAGCTGATG
TCTAGGATTTTTTTTACCCAAATGCTGGCTGAATGCTGTGAAATAAATTAAG
GTTTTGAAGAAAGTGTTCATGAAT

132 **atg**gagcgtctgaatacagcagctgcacaagcctcgtct
M E R L N T A A A Q A S S
171 tcgctttatggagttaacatggcagaatatggagatgctggcgtc
S L Y G V N M A E Y G D A G V
216 agctcgatgatggggttgataaccaacatgagcctgagcaaagc
S S M M G L I T Q H E P E Q S
261 gaaaatattatgtccacgaggatcccaagctcgttttcatcattc
E N I M S T R I P S S F S S F
306 catggccatgccgattgccttctctcagcagcaatgtttcagggt
H G H A D C L L S A A M F Q G
351 tctcaaggagatcataagctcaatccacagcctgggatgaaccag
S Q G D H K L N P Q P G M N Q
396 cagctagtctctgagcagcttatcatgtcagattcgtccatgccg
Q L V S E Q S I M S D S S M P
441 ttgggttaagacaaaagcttgctctgggtcttcgtaatcagtttgaa
L V K T K A C S G L R N Q F E
486 tttcacaggaacaacccggaaattgctacacagatcagtcctca
F H R E Q P G N C Y T D Q S S
531 aatattccgctaagccccatagtcacatcgttagcctcgcaggct
N I P L S P I V T S L A S Q A
576 cgaggagaagcgcggatgataccgctccttgatgccaacagtgct
R G E A R M I P S L D A N S A
621 catttcaatgtggataacgaggagcatgcaataaaatcgaaaatc
H F N V D N E E H A I K S K I
666 ttagcgcacccacagtatccgagcttggtgggagcttacatcgac
L A H P Q Y P S L L G A Y I D
711 tgtcaaaagattggggctcctccagaagctgtggcccgtttgat
C Q K I G A P P E A V A R L D
756 gctctaactcgcgaacaccaagaccacacagctcggactgtgagc
A L T R E H Q D P Q R R T V S
801 atcggaatggaccacagaactcgatcaatattatggaggcatactgc
I G M D P E L D Q F M E A Y C
846 gagatattgactaagtaccatgaagagcttgccaaacctttcaa
E I L T K Y H E E L A K P F K
891 gaagcaatgctgtttttgaagaagattgaaacccagttcaactcc
E A M L F L K K I E T Q F N S
936 ctgggaaaaggaacgataagaatcttctccaggaataactcgc
L G K G T I R I S S P G N T R
981 ttgcttcataaacatTTTTTacttctTTTTTctcctattgctcca
L L H K H F L L L F S P I A P
1026 attaatctgcttcatatcaactccactatgtaaattcacgttat
I N S A S Y Q L H Y V N S R Y
1071 tcgggagataaaaact**tga** 1088
S G D K T *

AGGAAATCAATGGGCTTCGAGAAAATCCAGTGAAAACCGAACTAGCTGTAAT
TGCGTTAAGTACAAGTCTAGACTGCCAAATGTAGATTATTAATGTTGAGGAC
AAAAATCTCAGGGCCATGGACGATATCGAGGGATGGAATTTAAATGAACTAT
TTGTTCTTATTATTTTTGGTGTCTATATGCCACTATAATGGCAAGACCTCAA
ATCATTATGGATATCATCAAGCCATGAAAGTCTTAAGGGGGTTTACAGAAAAC
CGCTGGGGGGTTACTGTTACTATTTCATCACTACTTTAAATATAAACCCAGTAC
TAAATATCCAGGTCTGAAGGGGCTTACAGGAAACCGCTGGGAGGTTAGTGTTTCG
TATTCATCGCTATTTTTAAATATAAACACAGTACTAAATATTCACGTCCTACAT
TAAGCATTATTTTTGGTGTCAATTAATGGAGATATAGTTGATAGCTTTTCATT
AGTGAGTGGCACTTATGGTTGTAGTACGTACAGCTAGTTGTAGACTGCTTTAT

ATAGTCTAAGATGACATTCTAGTCTTAGGCATATGACTTGTTGTAATGCCAG
ACTTATTATACC

Figure A.2: Nucleotide and deduced amino acid sequence of PtKN1(hd-).

The beginning of the sequence is the 5' end of PtKN1 cDNA. The deduced amino acid sequence is shown below the nucleotide sequence. The start and stop are highlighted in bold. The end of the sequence is the end of the cDNA.

Pt KN1 (hd-) TCA CCCAAATGCTGGCTGAAT GCT GTGAAA TAAATT AAA GGT TTT GAA GAAAGT GTTTCA
Pt KN1 (HD+) TCA CCCAAATGCTGGCTGAAT GCT GTGAAA TAAATT AAA GGT TTT GAA GAAAGT GTTTCA
consensus TCA CCCAAATGCTGGCTGAAT GCT GTGAAA TAAATT AAA GGT TTT GAA GAAAGT GTTTCA

Pt KN1 (hd-) TGAATATGGAGCGTCTGAATA CAGCAGCTGCACAAGCCT CGT CTT CGCTTT ATGGAGTTA
Pt KN1 (HD+) TGAATATGGAGCGTCTGAATA CAGCAGCTGCACAAGCCT CGT CTT CGCTTT ATGGAGTTA
consensus TGAATATGGAGCGTCTGAATA CAGCAGCTGCACAAGCCT CGT CTT CGCTTT ATGGAGTTA

Pt KN1 (hd-) ACATGGCAGAATATGGAGATGCTGGCGTCA GCT CGATGATGGGGT TGA TAA CCCAACATG
Pt KN1 (HD+) ACATGGCAGAATATGGAGATGCTGGCGTCA GCT CGATGATGGGGT TGA TAA CCCAACATG
consensus ACATGGCAGAATATGGAGATGCTGGCGTCA GCT CGATGATGGGGT TGA TAA CCCAACATG

Pt KN1 (hd-) AGCCTGAGCAAA GCGAAAATATTA TGT CCA CGA GGA TCCCAA GCT CGT TTT CAT CATTCC
Pt KN1 (HD+) AGCCTGAGCAAA GCGAAAATATTA TGT CCA CGA GGA TCCCAA GCT CGT TTT CAT CATTCC
consensus AGCCTGAGCAAA GCGAAAATATTA TGT CCA CGA GGA TCCCAA GCT CGT TTT CAT CATTCC

Pt KN1 (hd-) ATGGCCATGCCGATT GCCTTCTCT CAGCAGCAATGT TTCAGGGT CTCAAGGAGATCATA
Pt KN1 (HD+) ATGGCCATGCCGATT GCCTTCTCT CAGCAGCAATGT TTCAGGGT CTCAAGGAGATCATA
consensus ATGGCCATGCCGATT GCCTTCTCT CAGCAGCAATGT TTCAGGGT CTCAAGGAGATCATA

Pt KN1 (hd-) AGCTCAATCCACAGCCTGGGATGA ACCAGCAGCTAGTCT CTGAGCAGTCTATCATGT CAG
Pt KN1 (HD+) AGCTCAATCCACAGCCTGGGATGA ACCAGCAGCTAGTCT CTGAGCAGTCTATCATGT CAG
consensus AGCTCAATCCACAGCCTGGGATGA ACCAGCAGCTAGTCT CTGAGCAGTCTATCATGT CAG

Pt KN1 (hd-) ATT CGT CCATGCCGT TGGTTAAGA CAAAAGCTT GCT CTGGTCTTCGTAATCAGT TTTGAAT
Pt KN1 (HD+) ATT CGT CCATGCCGT TGGTTAAGA CAAAAGCTT GCT CTGGTCTTCGTAATCAGT TTTGAAT
consensus ATT CGT CCATGCCGT TGGTTAAGA CAAAAGCTT GCT CTGGTCTTCGTAATCAGT TTTGAAT

Pt KN1 (hd-) TTCACAGGGAACAACCCGAAATT GCT ACA CAGATCAGT CCT CAAATA TCCGCTAAGCC
Pt KN1 (HD+) TTCACAGGGAACAACCCGAAATT GCT ACA CAGATCAGT CCT CAAATA TCCGCTAAGCC
consensus TTCACAGGGAACAACCCGAAATT GCT ACA CAGATCAGT CCT CAAATA TCCGCTAAGCC

Pt KN1 (hd-) CCATAGTCA CAT CGT TAGCCT CGCAGGCTCGAGGAGAAGCGCGGA TGA TACCCT CCTTGG
Pt KN1 (HD+) CCATAGTCA CAT CGT TAGCCT CGCAGGCTCGAGGAGAAGCGCGGA TGA TACCCT CCTTGG
consensus CCATAGTCA CAT CGT TAGCCT CGCAGGCTCGAGGAGAAGCGCGGA TGA TACCCT CCTTGG

Pt KN1 (hd-) ATGCCAACAGTGCTCATT TCAATGTGGATAACGAGGAGCATGCAA TAAAT CGAAAA TCT
Pt KN1 (HD+) ATGCCAACAGTGCTCATT TCAATGTGGATAACGAGGAGCATGCAA TAAAT CGAAAA TCT
consensus ATGCCAACAGTGCTCATT TCAATGTGGATAACGAGGAGCATGCAA TAAAT CGAAAA TCT

Pt KN1 (hd-) TAGCGCACCCACAGTATCCGAGCT TGT TGGGAGCTT ACATCGACT GTCAAAAGATTGGGG
Pt KN1 (HD+) TAGCGCACCCACAGTATCCGAGCT TGT TGGGAGCTT ACATCGACT GTCAAAAGATTGGGG
consensus TAGCGCACCCACAGTATCCGAGCT TGT TGGGAGCTT ACATCGACT GTCAAAAGATTGGGG

Pt KN1 (hd-) CTCCTCCAGAAGCTGTGGCCCGTT TGGATGCTCTAA CTGCGAACACCAAGACCCACAAC
Pt KN1 (HD+) CTCCTCCAGAAGCTGTGGCCCGTT TGGATGCTCTAA CTGCGAACACCAAGACCCACAAC
consensus CTCCTCCAGAAGCTGTGGCCCGTT TGGATGCTCTAA CTGCGAACACCAAGACCCACAAC

Pt KN1 (hd-) GTCGGACTGTGAGCATCGGAATGGACCCAGA ACTCGATCAAT TTA TGGAGGCAT ACT GCG
Pt KN1 (HD+) GTCGGACTGTGAGCATCGGAATGGACCCAGA ACTCGATCAAT TTA TGGAGGCAT ACT GCG
consensus GTCGGACTGTGAGCATCGGAATGGACCCAGA ACTCGATCAAT TTA TGGAGGCAT ACT GCG

Pt KN1 (hd-) AGATATTGACTAAGT ACCATGAAGAGCTTGCCAAACCTT TCAAAGAAGCAA TGCTGTTTT
Pt KN1 (HD+) AGATATTGACTAAGT ACCATGAAGAGCTTGCCAAACCTT TCAAAGAAGCAA TGCTGTTTT
consensus AGATATTGACTAAGT ACCATGAAGAGCTTGCCAAACCTT TCAAAGAAGCAA TGCTGTTTT

Pt KN1 (hd-) TGAAGAAGATTGAAA CCCAGT TCAACT CCCTGGGAAAAGGAA CGA TAA GAA TTT CTT CTC
Pt KN1 (HD+) TGAAGAAGATTGAAA CCCAGT TCAACT CCCTGGGAAAAGGAA CGA TAA GAA TTT CTT CTC
consensus TGAAGAAGATTGAAA CCCAGT TCAACT CCCTGGGAAAAGGAA CGA TAA GAA TTT CTT CTC

Pt KN1 (hd-) CAGGTAATACTCGCTTGCTTCATAAACATT TTTTACTTCT TTTTTCTCCTATTGCTCCAA
Pt KN1 (HD+) CAGCTGACGATGACGAAAAGA CCGAGGAGGTGGATCTT CCGAGGAGGTCCAGGAGGCCA
consensus CAG- TaAt a- T- gC- -g- - - - - aAa- a- - - T- - Act Tct - - - - - ct a- - G- t- - - A

Pt KN1 (hd-) TTAATTCTGCTT CATATCAACTCCACTATGTAAT T CACGTTATT CCGGAGAT AAACTT
Pt KN1 (HD+) GT- GGTGGT GAAACGGATTTT CAGGAAGTGGAT CACCATGCAGTAGAGGATAGGGAGTTA
consensus - T- a- T- - - - - - a- c- - ct- - a- - aTG- A- - - t CAc Gt - aT- - gGGA- A- aaAacT-

Pt KN1 (hd-) GAAGGAAATCAATGGGCTTCGAGAAAATCCAGTGA AACCGAACTAGCTGTAAATTGCGT
Pt KN1 (HD+) AAAAAATCATCTCCTT CGCAAAATACAGTGGATATTTGAGTAGTCTAAAACAGGAATTCATG
consensus gAAGg- - ATC- - t- - - - t- - g- g- Aa- - - - gT- - aAc- G- - c- Ag- tg- aAATT- - -

```

Pt KN1 ( hd- ) TAAGTACAACTCTAGACTGCCAAATGTAGATTATTAATGTTGAGGACAAAATCTCAGG
Pt KN1 ( HD+ ) AAGAAGAAAAGAAAAGAAAGCTCCCTAAAATGCAAGGCAAAGTTACTCGATGGTGG
consensus - Aag- a- AAg- - - Ag a- - g- - - - - t - TAgA- - - - t A- a- - - - gAG- - - - - aATc- - - GG

Pt KN1 ( hd- ) GCCATGGACGATATCGAGGGATGGAATTTAAATGAAATTATTTGTTCTTATTATTTTTGG
Pt KN1 ( HD+ ) AGTCTGCACGACAAGTGGCCTTATCCTTCGCAGACGCAGAAAATAGCTTTGGCTGAATGC
consensus g- c- TG- ACGAt A- - - aG- - - Tg- - - TTt aaA- g- aa- - A- - - - - CTT- - - - T- - - TG-

Pt KN1 ( hd- ) TGTCTATATGCCACTATAATGGCAAGACCTCAATCATTATGGATATCATCAAGCCATG
Pt KN1 ( HD+ ) ACGGGGTTGGATCAAAAACAAATAAATAATTTGGTTTATAAACCAAAGAAAACGCCACTGG
consensus - - - - - Ta- gc CA- - A- A- g- - AA- A- c T- a- - T- - - - At - gA- a- - A- Ca- - - C- G

Pt KN1 ( hd- ) AAAGTCTTAAGGGGGTTTACAGAAAACCGCTGGGGGCTTACTGTTACTATTCATCACTAC
Pt KN1 ( HD+ ) AAGCCTTCTGAAGATATGCACTTCATGGTAATGAAAGTCAAGTCTCAGAGTGCAGCC
consensus AAa- t c Tt - agg Gg- - T- - - - - A- - - - - Ggg- g- T- - t g- T- CT- - t - aT- - c- - C

Pt KN1 ( hd- ) TTTAAATATAAACCCAGTACTAAATATCCAGGTCCAAGGGGCTTACAGGAAACCGCTGGG
Pt KN1 ( HD+ ) TTGTATGTTGAGAGACATATGATGACTGAAAGGTATCTTTAGATTGCTAGAAAGAACCTT
consensus TT- - A- - - TaAa- - - - gTAc- A- a- - T- - AGG- cg- - - - - g- - T- - - - gaAA- - g- t- - -

Pt KN1 ( hd- ) AGGTTAGTGTTCGTATTCATCGCTATTTTAAATATAACACAGTACTAATATTCAAGTC
Pt KN1 ( HD+ ) CGGCTGAAAACAGCACACAATGCTATTGCTTTTGTGATTTAATGGCATGGCTTCAA
consensus - GGt Tag- g- t - Gt At - CA- c GCTATT- t- - - TaT- a- - - c- g- - c- a- ATa- t c- c- -

Pt KN1 ( hd- ) CTACATTAAGCATTATTATTTGGTGTCAATTAATGGAGATATAGTTGATAGCTTTCATTA
Pt KN1 ( HD+ ) TTTC-----
consensus cT- C-----

Pt KN1 ( hd- ) GTGAGTGGCACTTATGGTGTAGTACGTACAGCTAGTTGTAGACTGCTTTATATAGTCTA
Pt KN1 ( HD+ ) -----
consensus -----

Pt KN1 ( hd- ) AGATGACATTCTAGTCTTAGGCATATGACTTGTGTAAATGCCAGACTTATTATACC
Pt KN1 ( HD+ ) -----
consensus -----

```

Figure A.3: Alignment of the nucleotide sequences of PtKN1(HD+) and PtKN1(hd-). In the region that is common to both mRNA forms, there is overall amino acid identity of 100% between the sequences.

```

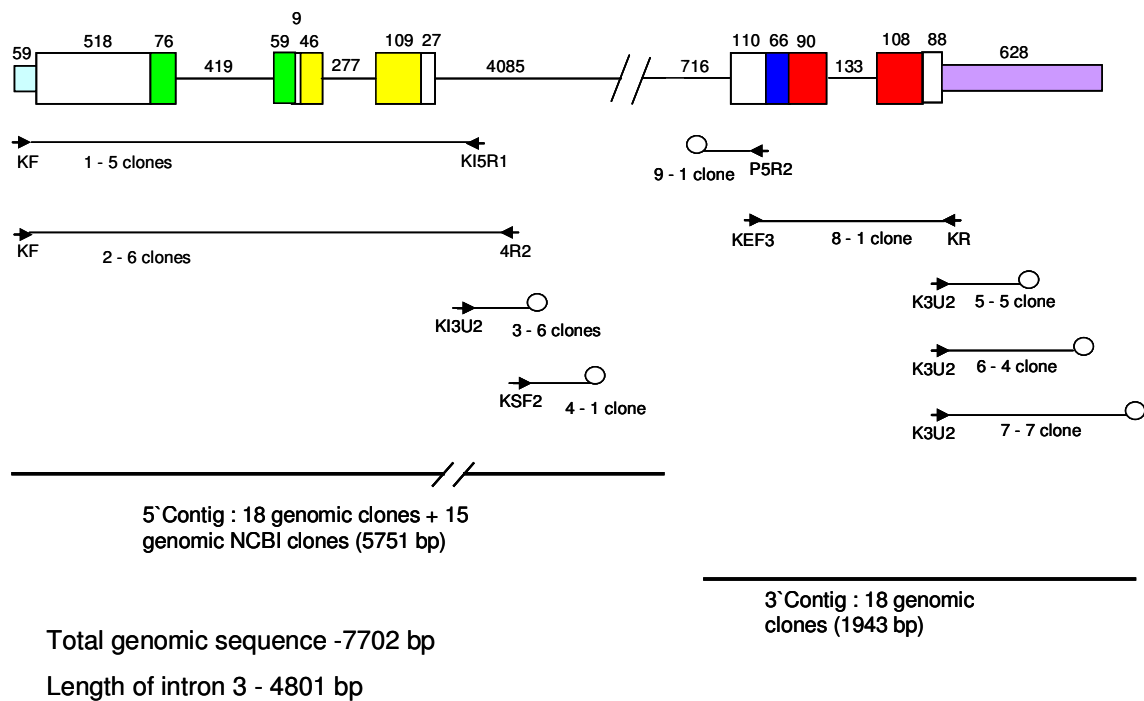
>PtKN1 (-)
MNMERLNTAAAQASSSLYGVNMAEYGDAGVSSMMGLITQHEPEQSENIMSTRIPSS
FSSFHGHADCLLSAAMFQGSQGDHKLNPQPGMNOQLVSEQSIMSDSSMPLVKTAC
SGLRNQFEFHREQPGNCYTDQSSNIPLSPIVTSLASQARGEARMIPSLDANSAHFN
VDNEEHAIKSKILAHPOYPSELLGAYIDCQKIGAPPEAVARLDALTRHQDPQRRTV
SIGMDPELDQFMEAYCEILTKYHEELAKPFKEAMLFLKKIETQFNLSLKGKTIRISS
PGNTRLLHKHFLLLFSPAPIPINSASYQLHYVNSRYS GDKT

>PtKN1 (+)
MNMERLNTAAAQASSSLYGVNMAEYGDAGVSSMMGLITQHEPEQSENIMSTRIPSS
FSSFHGHADCLLSAAMFQGSQGDHKLNPQPGMNOQLVSEQSIMSDSSMPLVKTAC
SGLRNQFEFHREQPGNCYTDQSSNIPLSPIVTSLASQARGEARMIPSLDANSAHFN
VDNEEHAIKSKILAHPOYPSELLGAYIDCQKIGAPPEAVARLDALTRHQDPQRRTV
SIGMDPELDQFMEAYCEILTKYHEELAKPFKEAMLFLKKIETQFNLSLKGKTIRISS
PADDDEKTEGGGSSEEVEDGSGGETDFQEVDDHHAIVEDRELKNHLLRKYSGLSLLK
QEFMKKKKKKGKLPKDARQKLLDWWSLHDKWPYPSETEKIALAECTGLDQKQINNWF
INQRKRHWKPS E DMHFMVMNSHSPHSAALYVERHMMTEGYL

```

Figure A.4: Peptide sequence used for making antibody for PtKN1(HD+) and PtKN1(hd-) proteins.

Region in green is KNOX1 domain, in yellow is KNOX2, in blue is ELK domain and in red is HD. The sequence highlighted in orange in both forms is the peptide sequence against which antibody was raised to perform western blotting.



1 = KG1, KG2, KG24, KG25

2 = KG3, KG5, KG6, KG7, KG9, KG10

3 = K415, K416, K417, K421, K424, K428

NCBI genomic clones used are gill62314188|gb|ET183639.1|ET183639, gill62314423|gb|ET183775.1|ET183775, gil75689884|gb|CZ895398.1|CZ895398, gill62313798|gb|ET184558.1|ET184558, gil62313648|gb|ET182707.1|ET182707, gnlltil2072254109, gnlltil2072254128, gil62315122|gb|ET182944.1|ET182944, gil75690342|gb|CZ895856.1|CZ895856, gil62313302|gb|ET182445.1|ET182445, gil75690579|gb|CZ896093.1|CZ896093, gil75689785|gb|CZ895299.1|CZ895299, gil62315797|gb|ET184334.1|ET184334, gil62314788|gb|ET183942.1|ET183942, gil62314827|gb|ET183958.1|ET183958

5' Contig = Genomic clones in 1 + 2 + 3 + following NCBI genomic clones

5 = G11, G13, G14, G15, G16

6 = G22, G23, G25, G26

7 = G31, G32, G33, G34, G35, G36, G37

8 = KG3

9 = KGW

3' Contig = Genomic clones in 5 + 6 + 7 + 8 + 9

Figure A.5: Genomic sequence map.

Using either 2 gene specific primers in PCR or 1 gene specific primer and adapter primer in genomic walk, I isolated several overlapping clones different regions of PtKN1 gene. Several genomic clone sequences from NCBI database overlapping some regions were found and contigs were made; a total of 7.7kb sequence of the PtKN1(HD+/hd-) gene was obtained. KNOX1; yellow - KNOX2; blue - ELK; red - HD; turquoise - 5' UTR; purple - 3' UTR of (HD+); open circle - adapters from the genome walker library.

```

NM_116884.3 CCTTGACGAATCTATATACCTAGTT CGTTTTTCTTCCTCAAATATATCTTTTTCAATT
KNAT1[hd-] -----
consensus -----
NM_116884.3 TATTTGGTTTTCTTTGGGTGCAACTTCACTCACAAAATTTCTCTCTTTTTTATATT
KNAT1[hd-] -----
consensus -----
NM_116884.3 AATTTGAGTTAGGCCTTTT GATTT CATAGAT GAGTCGTCTAGT CGTCTGGATTTGATGT
KNAT1[hd-] ----- GAGTCGTCTAGT CGTCTGGATTTGATGT
consensus ----- GAGTCGTCTAGT CGTCTGGATTTGATGT
NM_116884.3 GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAA GAATACATACATC
KNAT1[hd-] GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAA GAATACATACATC
consensus GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAA GAATACATACATC
NM_116884.3 CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
KNAT1[hd-] CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
consensus CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
NM_116884.3 TTTTTTCTTTGATTTTTATATTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
KNAT1[hd-] TTTTTTCTTTGATTTTTATATTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
consensus TTTTTTCTTTGATTTTTATATTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
NM_116884.3 AAGAATACCAGCATGACAACAGCACCCTCTCAAAGTAAGTTTCTTGACTCTCCAA
KNAT1[hd-] AAGAATACCAGCATGACAACAGCACCCTCTCAAAGA-----
consensus AAGAATACCAGCATGACAACAGCACCCTCTCAAAGA-----
NM_116884.3 TCTCTTCTCCAACA AAAACGATAACACAAGTGATACCAACAACAACAACAATAATA
KNAT1[hd-] -----
consensus -----
NM_116884.3 ATAGTAGCAATTATGGTCTGGTTACAATAACTAACACAACAATCATCACCACCAAC
KNAT1[hd-] -----
consensus -----
NM_116884.3 ACATGTTGTTTCCACATATGAGCTCTCTCTCCCTCAAACAACCGAGAATTGCTCCGAT
KNAT1[hd-] -----
consensus -----
NM_116884.3 CTGATCATGATCAACCCAACAACAACAACCCATCTGTTAAATCTGAAGCTAGCTCCT
KNAT1[hd-] -----
consensus -----
NM_116884.3 CAAGAATCAATCATTA CTCCATGTTAATGAGAGCCATCCACAATACTCAAGAGCTAACA
KNAT1[hd-] -----
consensus -----
NM_116884.3 ACAACAACAATGACAACGTAAGCGATGTTGAAGCCATGAAGGCTAAAATCATTGCTCATC
KNAT1[hd-] -----
consensus -----
NM_116884.3 CTCACTACTCTACCCTCTCAAGCTTACTTGGACTGCCAAAAGATTGGAGCTCCACCTG
KNAT1[hd-] -----
consensus -----
NM_116884.3 ATGTGGTTGATAGAATTACGGCGGCA CGGCAAGACTTTGAGGCTCGACAA CAGCGGTCAA
KNAT1[hd-] -----
consensus -----
NM_116884.3 CACCGTC- TGTCTCTGCCTCCTCTAGAGACCCGGAGTTAGATCAATT CATGGAAGCATAAC
KNAT1[hd-] ----- GAAGCATAAC
consensus ----- GAAGCATAAC
NM_116884.3 TGTGACATGTTGGTTAAATATCGTGAAGGCTAACAAGGCCCATTCAGGAAGCAATGGAG
KNAT1[hd-] TGTGACATGTTGGTTAAATATCGTGAAGGCTAACAAGGCCCATTCAGGAAGCAATGGAG
consensus TGTGACATGTTGGTTAAATATCGTGAAGGCTAACAAGGCCCATTCAGGAAGCAATGGAG
NM_116884.3 TTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC
KNAT1[hd-] TTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC
consensus TTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC

```

Figure A.6: Alignment of the nucleotide sequences of *Arabidopsis* KNAT1(hd-) and KNAT1 full length mRNA.


```

NC_003075.4 -----TAGAATACGGCGGCACGGCAAGACTTTGAGGCTCGACAACAGCGGTCA
KNAT1[hd-]
NM_116884.3 GATGTGGTTGATAGAAATACGGCGGCACGGCAAGACTTTGAGGCTCGACAACAGCGGTCA
consensus -----t agaat t acggcggc acggcaagact t t gaggct cgacaacagcggtc a

NC_003075.4 ACACCGTCTGTCTCTGCCTCCTCTAGAGACCCGGAGTTAGATCAATTCAT-----
KNAT1[hd-] -----CCCGGAGTTAGATCAATTCAT-----
NM_116884.3 ACACCGTCTGTCTCTGCCTCCTCTAGAGACCCGGAGTTAGATCAATTCATGGTTCATTAC
consensus acaccgtctgtctctgcctcctctagagaCCCGGAGTTAGATCAATTCAT-----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 TAACCAAAGTTGAATTTAATTAATGGGTCGGTCTTTGTGTGTATATGTGTAGAAGT
consensus -----

NC_003075.4 -----GGAAGCATACTGTGACATGTTGGTTA
KNAT1[hd-] -----GGAAGCATACTGTGACATGTTGGTTA
NM_116884.3 CTTGATTTTCCTTTTTATTTTTTGGTATGAATAGGAAGCATACTGTGACATGTTGGTTA
consensus -----GGAAGCATACTGTGACATGTTGGTTA

NC_003075.4 AATATCGTGAGGAGCTAACAAGGCCATT CAGGAA GCAATGGAGTTTATACGTCGTATTG
KNAT1[hd-] AATATCGTGAGGAGCTAACAAGGCCATT CAGGAA GCAATGGAGTTTATAC-----
NM_116884.3 AATATCGTGAGGAGCTAACAAGGCCATT CAGGAA GCAATGGAGTTTATACGTCGTATTG
consensus AATATCGTGAGGAGCTAACAAGGCCATT CAGGAA GCAATGGAGTTTATACGt cgt at t g

NC_003075.4 AATCTCAGCTTAGCATGTTGTGT CAGAGTCCCATTCACATCCTCAACAATCCTG-----
KNAT1[hd-] -----
NM_116884.3 AATCTCAGCTTAGCATGTTGTGT CAGAGTCCCATTCACATCCTCAACAATCCTGACAT
consensus aatctcagcttagcatgttgtgt cagagtcccattcacatcctcaacaatcctg-----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 GTCATCAATCTCATATACATACATATATGCATGTAA GAAATGGCTGGAGAAAACAAAG
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 TGTAGTGGCAAGGTAGGCTAATGATGGATTGCATAAATTGGCATGTTCTTTGGTTCAT
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 GGTA AATTGTTGTAACATTGAGATATTA AAAAAAAAAAAAAATCATAGTCTACATACATT
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 TCAAAATGTTAAATAAGATTATAAGATAATTAAGATAATTGAGATGCATTATTGTTGGA A
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 GTATGTAAGATACAAATAGGTGTGCATATTTGTTCTTGTTT TCCCATTTGGGAATCCAT
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 CCAAA GATGGGTGGGGCTATTTTGT TTTTCTTTTCACTCGGCCAAAAGTACTAT
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 CACAAGAACTTTGGTGTCTGCTGAAGATTATTTGGACATATCTTCTTCGATTT C
consensus -----

NC_003075.4 -----
KNAT1[hd-]
NM_116884.3 GTTATGCTAAAACATCAAAATCTCTACTCGATTAATTAGTCTAATCATAAAGGTAGATG
consensus -----

```

Figure A.7: Alignment of the nucleotide sequences of *Arabidopsis* KNAT1(hd-), KNAT1 full length mRNA and KNAT1 gene. (page 1)

```

NC_003075.4 -----
KNAT1[hd-] 3 ACTAGAATGTTGTTAAATGACTAAACCATTTAACTTCATGAAATATATTATGTTGAAGCA
NM_116884.3 consensus -----
NC_003075.4 - ATGGGAAAGAGTGACAAATATGGGATCATCAGACGAAGAACAAGAGAATAACAGCGGAGGG
KNAT1[hd-] 3
NM_116884.3 GATGGGAAAGAGTGACAAATATGGGATCATCAGACGAAGAACAAGAGAATAACAGCGGAGGG
consensus -atgggaagagt gacaat at ggg at cat c agacgaagaacaagagaat aacagcggaggg
NC_003075.4 GAAACAGAATTACCGGAAATAGACCCGAGGGCCGAAGATCGGGAACTCAAGAACCATTTG
KNAT1[hd-] 3
NM_116884.3 GAAACAGAATTACCGGAAATAGACCCGAGGGCCGAAGATCGGGAACTCAAGAACCATTTG
consensus gaaacagaat t accggaat agacccgagggccgaagatcgggaact caagaaccatt t g
NC_003075.4 CTGAAGAAGTATAGTGGATACTTAAGCAGTTTGAAGCAAGAAGTATCCAAGAAAGAAAAG
KNAT1[hd-] 3
NM_116884.3 CTGAAGAAGTATAGTGGATACTTAAGCAGTTTGAAGCAAGAAGTATCCAAGAAAGAAAAG
consensus ct gaagaagt at agtggat act t aagcagttt gaagcaagaact atccaagaagaaaag
NC_003075.4 AAAGGTAAACTTCTAAAGAAGCACGGCCGAAAGCTTCTCACGTGGTGGGAGTTGCATTAC
KNAT1[hd-] 3
NM_116884.3 AAAGGTAAACTTCTAAAGAAGCACGGCCGAAAGCTTCTCACGTGGTGGGAGTTGCATTAC
consensus aaaggt aaact t cct aaagaagc acggcagaagct t ct cacgt ggt gggagt t gcat t ac
NC_003075.4 AAGTGGCCATATCCTTCT-----
KNAT1[hd-] 3
NM_116884.3 AAGTGGCCATATCCTTCTGTACGTATAAATTTTCTCTCATCTCTTTTGCCTCATATTT
consensus aagtggccat at cct t ct-----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 TTTAAATATATACTACAAATTAAGTGTTTAAAAACCTATCTTAAGCTAGCAATCCC
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 TCTGAAGCCGCTATGTTAGGCCAAGTAGGTACTTGCATGACCGATTTTGTGTTAAATCC
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 TTGTACGGTTTCCAACAACATGGTTTTAAATCCGAATGCTCACAACACTAGAATTAGTCA
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 CTAGGATACATATCTTAGTCATGGTGTCTAACCCCTAACATACTATTTTCTAACACGTA
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 TATCAATCACTTTTTTTTAAAAAATGTAAGTTTGTGATGAAATTTTTATCCATATATTTT
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 GAAACAATTTAATTTTTTTGTAGTTTTAATTAGTTTGAATGAAATAACTAATCTAACC
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 GACTGTAATGATAGTGAATAAACAATGTTAATATGGTAAATATAAATGACATGTTGTT
consensus -----
NC_003075.4 -----
KNAT1[hd-] 3
NM_116884.3 AAAAAGATAATAGACAACATATAGT TAGTCGTTACTT GATGTCAAAGAAAAATGAAGTT
consensus -----

```

Figure A.7: Alignment of the nucleotide sequences of *Arabidopsis* KNAT1(hd-), KNAT1 full length mRNA and KNAT1 gene. (page 2)

```

NC_003075.4 -----
KNAT1[hd-] -----
NM_116884.3 AGTCATTTTACTTAGAGAATCGTCAAAAACAACAACTTTAGGTCATTTTCCATTTTAT
consensus -----

NC_003075.4 -----
KNAT1[hd-] -----
NM_116884.3 GTCTAGAATCATTTAAAAATAATTTTCGATTATCTCTAGCATTGCTTTTATACATTGAT
consensus -----

NC_003075.4 ----- GAGTCAGAGAAGGTAGC
KNAT1[hd-] -----
NM_116884.3 TACTACATAAATGAAGATGGTTGGTT--TGAGTTGTGTGGCAGGAGTCAGAGAAGGTAGC
consensus ----- g agt cag aga aggt agc

NC_003075.4 GTTGGCGGAATCAACGGGGTTAGATCAGAAACAAATCAACAATTTGGTTCATAAACCAAAG
KNAT1[hd-] -----
NM_116884.3 GTTGGCGGAATCAACGGGGTTAGATCAGAAACAAATCAACAATTTGGTTCATAAACCAAAG
consensus gt t ggc gga at caac ggg g t t ag at c aga aac aa at ca ca at t ggt t c at aa acc aa g

NC_003075.4 -- AAAGCGTCACTGGAAACCATCTGAAGACATGC- AGTTCATGGTGATGGATGGTCTGCA
KNAT1[hd-] -----
NM_116884.3 -- AAAGCGTCACTGGAAACCATCTGAAGACATGC- AGTTCATGGTGATGGATGGTCTGCA
consensus -- aaagcgt cact ggaa acc at ct ga ag ac at gc- agt t cat ggt gat gg at ggt ct gc a

NC_003075.4 GCACCCGCACCACGCAGCTCTGTACATGGATG GTCATTACATGGGTGATGGACCTTATCG
KNAT1[hd-] -----
NM_116884.3 GCACCCGCACCACGCAGCTCTGTACATGGATG GTCATTACATGGGTGATGGACCTTATCG
consensus gc acc cgc acc acg cag ct ct gt ac at gg at g GTCATTACATGGGTGATGGACCTTATCG

NC_003075.4 TCTCGGTCCATAAGACATCCAAAAGCTTTAGCCACATAATAACAACCTCTCGTTGCTTTTC
KNAT1[hd-] -----
NM_116884.3 TCTCGGTCCATAAGACATCCAAAAGCTTTAGCCACATAATAACAACCTCTCGTTGCTTTTC
consensus TCTCGGTCCATAAGACATCCAAAAGCTTTAGCCACATAATAACAACCTCTCGTTGCTTTTC

NC_003075.4 TTGTTACAACCTCATGTTTTGAATCCCCTACATCAGTTTGCTACTTATAGCTTTCTTTGT
KNAT1[hd-] -----
NM_116884.3 TTGTTACAACCTCATGTTTTGAATCCCCTACATCAGTTTGCTACTTATAGCTTTCTTTGT
consensus TTGTTACAACCTCATGTTTTGAATCCCCTACATCAGTTTGCTACTTATAGCTTTCTTTGT

NC_003075.4 TTTACGCCTTTTGTAATGTCTTATGTCGTTCCGGGAGTTTGAGACTTCCTAGTCAGAAA
KNAT1[hd-] -----
NM_116884.3 TTTACGCCTTTTGTAATGTCTTATGTCGTTCCGGGAGTTTGAGACTTCCTAGTCAGAAA
consensus TTTACGCCTTTTGTAATGTCTTATGTCGTTCCGGGAGTTTGAGACTTCCTAGTCAGAAA

NC_003075.4 TATCCCTTCATTTTATTTCTCTTTTTCGGTGTATTTGTTTTTTGTTTTCTCACTA
KNAT1[hd-] -----
NM_116884.3 TATCCCTTCATTTTATTTCTCTTTTTCGGTGTATTTGTTTTTTGTTTTCTCACTA
consensus TATCCCTTCATTTTATTTCTCTTTTTCGGTGTATTTGTTTTTTGTTTTCTCACTA

NC_003075.4 ATGTTTTTATTTAATAATGTGTGAAAGGAAAATGCTATTCTG-----
KNAT1[hd-] -----
NM_116884.3 ATGTTTTTATTTAATAATGTGTGAAAGGAAAATGCTATTCTG-----
consensus ATGTTTTTATTTAATAATGTGTGAAAGGAAAATGCTATTCTG-----

NC_003075.4 -----
KNAT1[hd-] -----
NM_116884.3 -----
consensus -----
TTTTTATTTTATATGGCATGAAATGTATGGTTTTT

```

Figure A.7: Alignment of the nucleotide sequences of *Arabidopsis* KNAT1(hd-), KNAT1 full length mRNA and KNAT1 gene. (page 3)

```

DY236300.1 -----
NP_001105331.1 MDGSF GNL GAGAGSSSGGSNSKAAAT AVSSSSFLQLPLSTASPAYYGAPLALLHHAAAAP
consensus mdqsf gnl gagagsssggsnskaa at avssssf l ql pl st asp ay ygapl al l hhaa ap

DY236300.1 -----
NP_001105331.1 SSSCQHCCQCHHHHYARHGAEMSAEAEAI KAKI VAHPCYSALLAAYLDCCKVGAPPDVL
consensus sssqqhqqqqhhhyar hgaems aaeaeai kaki vahpcysall aayl dcckvgappdvl

DY236300.1 -----
NP_001105331.1 ERLTAMAAKL DASAAGRHEPRDPEL DCFMEAYCNMLVKYREELTRPI DEAMEFLKRVEAQ
consensus erl t amaakl d asa agr hepr d pel dcf meaycnml vkyr eel tr pi deame fl kr veaq

DY236300.1 ----- EDDMDPNGRENDPPEI DPRAEDKELK-----
NP_001105331.1 LDCI SGGGGSSSARLSLADGKSEGVGSS EDDMDPNGRENDPPEI DPRAEDKELKYQLLKK
consensus l dci sggggsssar l sl adgksegvgss eddmdpngrendppe i dp raedkel kyql l kk

DY236300.1 ----- YPSETEKI ALAESTGLDCKQ
NP_001105331.1 YSGYLSSLRCEF SKKKKKGKLPKEARCKLLHWWELHYKWPYPSETEKI ALAESTGLDCKQ
consensus ysgyl ssl rcef skkkkkgkl pkearckll hwwel hykwp ypsete ki alaestgl dckq

DY236300.1 I NNWFFTQRKRHWKPS E DMP FVMVEGFHP CNA AAL YMDGPF NADGMYRLGS
NP_001105331.1 I NNWFI NQRKRHWKPS E DMP FVMVEGFHP CNA AAL YMDGPF NRDGMYRLGS
consensus I NNWF - - QRKRHWKPS E DMP FVMVEGFHP CNA AAL YMDGPF N - DGMYRLGS

```

Figure A.8: Alignment of the amino acid sequences of maize RS1 (NP_001105331.1) and deduced amino acid sequence of HD splice form mRNA (DY236300.1). Conserved domains are indicated as follows: KNOX1; yellow - KNOX2; blue - ELK; red - HD.

```

DY236300.1 -----
NM_001111861.1 CGGCGGCGGCGCT CCT CCT CAGCGCGCTCT CGCTCGCCGATGGTAAAT CAGAAGGGGT
consensus -----

DY236300.1 -----
NM_001111861.1 TGGCTCTTCTGAAGATGACATGGACCCGAATGGTCCGCGAGAACGATCCGCCGGAGATTGA
consensus -----
                    GAAGATGACATGGACCCGAATGGTCCGCGAGAACGATCCGCCGGAGATTGA

DY236300.1 -----
NM_001111861.1 CCCACGAGCGGAGGACAAGGAGCTCAA-----
consensus -----
                    CCCACGAGCGGAGGACAAGGAGCTCAAAGTACCAGCTGCTGAAGAAGTACAGTGGCTACCT
                    CCCACGAGCGGAGGACAAGGAGCTCAA-----

DY236300.1 -----
NM_001111861.1 GAGCAGCCTCAGGCAGGAGT TTT CCAAGAAAGAAGAAGAAAGGGAAAGCTCCCAAAGGAGGC
consensus -----

DY236300.1 -----
NM_001111861.1 CAGGCAGAAACTCCTCCTACTGGTGGGAGCTGCACTACAAGTGGCCGTACCCTTCAGAGAC
consensus -----
                    -----GTACCCTTCAGAGAC
                    -----GTACCCTTCAGAGAC

DY236300.1 -----
NM_001111861.1 GGAGAAGATTGCGCTCGCAGAATCAACGGGCCTAGACCAGAAGCAGATCAACAACCTGGTT
consensus -----
                    GGAGAAGATTGCGCTCGCAGAATCAACGGGCCTAGACCAGAAGCAGATCAACAACCTGGTT

DY236300.1 -----
NM_001111861.1 CTTCAACCAGAGGAAACGGCACTGGAAGCCATCGGAGGACATGCCTTTTCGTGATGATGGA
consensus -----
                    CTTCAACCAGAGGAAACGGCACTGGAAGCCATCGGAGGACATGCCTTTTCGTGATGATGGA
                    C-TCAACCAGAGGAAACGGCACTGGAAGCCATCGGAGGACATGCCTTTTCGTGATGATGGA

DY236300.1 -----
NM_001111861.1 AGGCTTCCACCCGAGAACGCCGCTGCTCTGTACATGGACGGGCCGTTTCATGGCCGACGG
consensus -----
                    AGGCTTCCACCCGAGAACGCCGCTGCTCTGTACATGGACGGGCCGTTTCATGGCCGACGG
                    AGGCTTCCACCCGAGAACGCCGCTGCTCTGTACATGGACGGGCCGTTTCATG- -CGACGG

DY236300.1 -----
NM_001111861.1 CATGTATCGCCTCGGTTTCATGAACTACACTTCCATGGAC- -GACTGTGGGGCCTATC
consensus -----
                    CATGTATCGCCTCGGTTTCATGAACTACACTTCCATGGAC- -GACTGTGGGGCCTATC
                    TATGTATCGCCTCGGTTTCATGAACTACACTTCCATGGACACGACTGTGGGG- CCTATC
                    cATGTATCGCCTCGGTTTCATGAACTACACTTCCATGGAC- -GACTGTGGGG- CCTATC

DY236300.1 -----
NM_001111861.1 CATTTCGTCCGCCCGTGGGAGTATATTCA-----
consensus -----
                    CATTTCGTCCGCCCGTGGGAGTATATTCA-----
                    CAATCGTCCGCCCGTGGGAGTATATTCA-----
                    CA-TCGTCCG- -C-GTGGGAGTATATTt- -a-----
                    -----GtGTTGCGGTGCAA

DY236300.1 -----
NM_001111861.1 GGACATTGTTGTGGAATTTTTGTGCTGATTGCCAAGGGGTGGTCCGTCGATGTAGCGT
consensus -----
                    GGACATTGTTGTGGAATTTTTGTGCTGATTGCCAAGGGGTGGTCCGTCGATGTAGCGT
                    GGACATTGTTGTGGAATTTTTGTGCTGATTGCCAAGGGGTGGTCCGTCGATGTAGCGT

DY236300.1 -----
NM_001111861.1 CTATGCATGGCTTGTAGTTCTCCTGTAAATGTTGTTGTTAATTTCCGTTGGCGCTTGT
consensus -----
                    CTATGCATGGCTTGTAGTTCTCCTGTAAATGTTGTTGTTAATTTCCGTTGGCGCTTGT
                    CTATGCATGGCTTGTAGTTCTCCTGTAAATGTTGTTGTTAATTTCCGTTGGC- - -TTGTA
                    CTATGCATGGCTTGTAGTTCTCCTGTAAATGTTGTTGTTAATTTCCGTTGGC- - -TTGTA

DY236300.1 -----
NM_001111861.1 ATCTAGGTGATTTCTGAACTTGAATCCTGGGTGATTAAATCCGCAAGCAGCTGCCTGT
consensus -----
                    ATCTAGGTGATTTCTGAACTTGAATCCTGGGTGATTAAATCCGCAAGCAGCTGCCTGT
                    ATCTAGGTGATTTCTGAACTTGAATCCTGGGTGATTAAATCCGCAAGCAGCTGCCTGT
                    ATCTAGGTGATTTCTGAACTTGAATCCTGGGTGATTAAATCCGCAAGCAGCTGCCTGT

DY236300.1 -----
NM_001111861.1 TTGTATGATTGTATGGTTCCGGTGGTGGAAACATGATGCATGCTAAGGCGTGGCCTGTCT
consensus -----
                    TTGTATGATTGTATGGTTCCGGTGGTGGAAACATGATGCATGCTAAGGCGTGGCCTGTCT
                    TTGTATGATTGTATGGTTCCGGTGGTGGAAACATGATGCATGCTAAGGCGTGGCCTGTCT

DY236300.1 -----
NM_001111861.1 GGATCAAAAAAAAAAAGATGTCTTTGTCCC
consensus -----
                    GGATC-----
                    GGATC-----

```

Figure A.9: Alignment of the nucleotide sequences of maize RS1(hd-) (DY23600.1) and RS1 full length mRNA (NM_001118.1).

```

KNAT[ knox- MEEYGHDNSTTPGR-----
NP_192555 MEEYGHDNSTTPGRVsflyspi sssnkndnt sdt nnnnnnnnssnygpgynnt nnnnhhh
consensus MEEYGHDNSTTPGRv s f l y s p i s s s n k n d n t s d t n n n n n n n n s s n y g p g y n n t n n n n h h h

KNAT[ knox- -----
NP_192555 CHMLFPHMSLLPCTTENCFRSDHDCPNNNNNPVSKSEASSSRI NHYSMLMRAI HNTCEA
consensus qhml f phr s s l l p q t t e n c f r s d h d c p n n n n n p s v k s e a s s s r i n h y s m l m r a i h n t q e a

KNAT[ knox- -----
NP_192555 NNNNNDVSDVEAMKAKI I AHPHYSTLLQAYLDCQKI GAPPDVVDRI TAARCDFEARCGR
consensus n n n n n d n v s d v e a m k a k i i a h p h y s t l l q a y l d c q k i g a p p d v v d r i t a a r q d f e a r q q r

KNAT[ knox- ----- EAYCDMLVKYREELTRPI CEAMEFI RRI ESCLSMCCSPI HI
NP_192555 STPSVSASSRDPELDGFMEAYCDMLVKYREELTRPI CEAMEFI RRI ESCLSMCCSPI HI
consensus s t p s v s a s s r d p e l d g f m e a y c d m l v k y r e e l t r p i c e a m e f i r r i e s c l s m c c s p i h i

KNAT[ knox- LNNPDGKSDNMGSSDEEQENSSGGETELPEI DPRAEDRELKNHLLKKYSGYLSSLKQELS
NP_192555 LNNPDGKSDNMGSSDEEQENSSGGETELPEI DPRAEDRELKNHLLKKYSGYLSSLKQELS
consensus l n n p d g k s d n m g s s d e e q e n s s g g e t e l p e i d p r a e d r e l k n h l l k k y s g y l s s l k q e l s

KNAT[ knox- KKKKKGKLPKEARQKLLTWELHYKVPYPSESEKVALAESTGLDQKQI NMF1 NGRKRHW
NP_192555 KKKKKGKLPKEARQKLLTWELHYKVPYPSESEKVALAESTGLDQKQI NMF1 NGRKRHW
consensus k k k k k g k l p k e a r q k l l t w e l h y k v p y p s e s e k v a l a e s t g l d q k q i n m f 1 n g r k r h w

KNAT[ knox- KPSEDMGFMVMDGLQHPHHAALYMDGHYMGDGPYRLGP
NP_192555 KPSEDMGFMVMDGLQHPHHAALYMDGHYMGDGPYRLGP
consensus k p s e d m g f m v m d g l q h p h h a a l y m d g h y m g d g p y r l g p

```

Figure A.10: Alignment of the amino acid sequences deduced from KNAT1(knox-) and *Arabidopsis* KNAT1 (NM_116884).

Conserved domains are indicated as follows: green - KNOX1; yellow - KNOX2; blue - ELK; red – HD.

```

NM_116884.3 CCTTGACGAATTCTATATACCTAGTT CGTTTTTTCTTCCTCAAATATATCTTTTCAATT
KNAT_KNOX-
consensus -----
NM_116884.3 TATTTGGTTTTCTTTGGGTGCAACTTCACCTCACAAAATTTCTCTCTTTTATATT
KNAT_KNOX-
consensus -----
NM_116884.3 AATTTGAGTTAGGCCTTTTGATTTCATAGATGAGTCGTCTAGTCGTCTGGATTTGATGT
KNAT_KNOX- GAGTCGTCTAGTCGTCTGGATTTGATGT
consensus ----- GAGTCGTCTAGTCGTCTGGATTTGATGT
NM_116884.3 GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAAGAATACATACATC
KNAT_KNOX- GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAAGAATACATACATC
consensus ----- GGTTATAGTCTTACAGAGACCTTTGATTGAAATAAGAA CAAAAGCAAGAATACATACATC
NM_116884.3 CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
KNAT_KNOX- CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
consensus ----- CTCTTCATCTTACACCCATCCTTTTTATTTTTCTAGGGTTTTATTTTTTTTTAATTTAT
NM_116884.3 TTTTTTCTTTGATTTTTATTTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
KNAT_KNOX- TTTTTTCTTTGATTTTTATTTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
consensus ----- TTTTTTCTTTGATTTTTATTTCTCTCTCTCTCTCAAATCTTTACTCATCTGGGTATGG
NM_116884.3 AAGAATACCAGCATGACAACAGCACCCTCTCAAAGTAAGTTTCTTGTACTCTCCAA
KNAT_KNOX- AAGAATACCAGCATGACAACAGCACCCTCTCAAAGA
consensus ----- AAGAATACCAGCATGACAACAGCACCCTCTCAAAGA
NM_116884.3 TCTCTCTCCAACAAAACGAT AACACAAGT GATACCAACAACAACAACAATATA
KNAT_KNOX-
consensus -----
NM_116884.3 ATAGTAGCAATTATGGTCTGGTTACAATAACTAACACAACAATCATCACCACCAAC
KNAT_KNOX-
consensus -----
NM_116884.3 ACATGTTGTTTCCACATATGAGCTCTCTCTCCCTCAAACAACCGAGAATTGCTTCCGAT
KNAT_KNOX-
consensus -----
NM_116884.3 CTGATCATGATCAACCCAACAACAACAACCCATCTGTTAAATCTGAAGCTAGCTCCT
KNAT_KNOX-
consensus -----
NM_116884.3 CAAGAATCAATCATTACTCCATGTTAATGAGAGCCATCCACAATACTCAAGAAGCTAACA
KNAT_KNOX-
consensus -----
NM_116884.3 ACAACAACAATGACAACGTAAGCGATGTTGAGGCCATGAAGGCTAAAATCATTGCTCATC
KNAT_KNOX-
consensus -----
NM_116884.3 CTCACTACTCTACCCTCTACAAGCTTACTTGGACTGCCAAAAGATTGGAGCTCCACCTG
KNAT_KNOX-
consensus -----
NM_116884.3 ATGTGTTGATAGAATTACGGCGGCA CGGCAAGACTTTGAGGCTCGACAA CAGCGGTCAA
KNAT_KNOX-
consensus -----
NM_116884.3 CACCGTC- TGTCTCTGCCTCCTCTAGAGACCCGGAGTTAGATCAATT CATGGAAGCATAAC
KNAT_KNOX- GAAGCATAAC
consensus ----- GAAGCATAAC
NM_116884.3 TGTGACATGTTGGTTAAATATCGTGAGGAGCTAACAAGGCCCATTCAGGAAGCAATGGAG
KNAT_KNOX- TGTGACATGTTGGTTAAATATCGTGAGGAGCTAACAAGGCCCATTCAGGAAGCAATGGAG
consensus ----- TGTGACATGTTGGTTAAATATCGTGAGGAGCTAACAAGGCCCATTCAGGAAGCAATGGAG
NM_116884.3 TTTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC
KNAT_KNOX- TTTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC
consensus ----- TTTTATACGTCGTATTGAATCTCAGCTTAGCATGTTGTGTCA GAGTCCCATTCACATCCTC

```

Figure A.11: Alignment of the nucleotide sequences of *Arabidopsis* KNAT1(knox-) and KNAT1 full length mRNA.

```

AK107637.1 -----
NM_001057674 -----
D16507.1 TTCTTTTTCTCTTTCTCTTCGTCCAAGAACTTCACTCAATAGCTCGAGCTACGGCT
consensus -----

AK107637.1 - AGCTCAGCTCAGCTCAAGGAGAGCTCGATCGCTGCACCAATACTTCACTGGAGATCGCC
NM_001057674 ----- AGGAGAGCTCGATCGCTGCACCAATACTTCACTGGAGATCGCC
D16507.1 AACTTTTTCGGTTCGGC- AGGAGAGCTCGATCGCTGCACCAATACTTCACTGGAGATCGCC
consensus - a-ctc-gc-c-gc-c- AGGAGAGCTCGATCGCTGCACCAATACTTCACTGGAGATCGCC

AK107637.1 TAGCTGCAGCTAGCTAGCTTATCCTGCGTGTCTTGAGCTTCTCTCATGGAGGAGATCTCC
NM_001057674 TAGCTGCAGCTAGCTAGCTTATCCTGCGTGTCTTGAGCTTCTCTCATGGAGGAGATCTCC
D16507.1 TAGCTGCAGCTAGCTAGCTTATCCTGCGTGTCTTGAGCTTCTCTCATGGAGGAGATCTCC
consensus TAGCTGCAGCTAGCTAGCTTATCCTGCGTGTCTTGAGCTTCTCTCATGGAGGAGATCTCC

AK107637.1 CACCACTTCGGGGTCGTGCGAGCCAGCGGCGTCCATGCGGCCACCAGCATCAGCATCAC
NM_001057674 CACCACTTCGGGGTCGTGCGAGCCAGCGGCGTCCATGCGGCCACCAGCATCAGCATCAC
D16507.1 CACCACTTCGGGGTCGTGCGAGCCAGCGGCGTCCATGCGGCCACCAGCATCAGCATCAC
consensus CACCACTTCGGGGTCGTGCGAGCCAGCGGCGTCCATGCGGCCACCAGCATCAGCATCAC

AK107637.1 CACCACCCGTGGGTTCTTCCCTGAGCGCCATCGTCGCGCCGCCCGCCGCCCGCAGCTG
NM_001057674 CACCACCCGTGGGTTCTTCCCTGAGCGCCATCGTCGCGCCGCCCGCCGCCCGCAGCTG
D16507.1 CACCACCCGTGGGTTCTTCCCTGAGCGCCATCGTCGCGCCGCCCGCCGCCCGCAGCTG
consensus CACCACCCGTGGGTTCTTCCCTGAGCGCCATCGTCGCGCCGCCCGCCGCCCGCAGCTG

AK107637.1 CAGCAGCAGCAGACGCAGGCCGCGGCTATGGCGCACACGCCGCTCACGCT-----
NM_001057674 CAGCAGCAGCAGACGCAGGCCGCGGCTATGGCGCACACGCCGCTCACGCT-----
D16507.1 CAGCAGCAGCAGACGCAGGCCGCGGCTATGGCGCACACGCCGCTCACGCTGAACACGGCG
consensus CAGCAGCAGCAGACGCAGGCCGCGGCTATGGCGCACACGCCGCTCACGCT-----

AK107637.1 -----
NM_001057674 -----
D16507.1 GCGGCGGCGGTGCGCAACCCGGTGCTGCAGCTGGCGAACGGCAGCCTGCTCGACGCGTGC
consensus -----

AK107637.1 -----
NM_001057674 -----
D16507.1 GGCAAGGCGAAGGAGGCGTGGCGTGGCGTCTACGCGCCGACGTTGGAGGCCATCAAG
consensus -----

AK107637.1 ----- ACCTCGACTGCCAG
NM_001057674 ----- ACCTCGACTGCCAG
D16507.1 GCCAAGATCATCTCCCACCCCACTACTCCTCCCTCCTCGCGCCTACCTCGACTGCCAG
consensus ----- ACCTCGACTGCCAG

AK107637.1 AAGGTTGGGGCGCCGCGGAGGTGGCGGCGAGGCTGACGGCGATGGCGCAGGACCTGGAG
NM_001057674 AAGGTTGGGGCGCCGCGGAGGTGGCGGCGAGGCTGACGGCGATGGCGCAGGACCTGGAG
D16507.1 AAGGTTGGGGCGCCGCGGAGGTGGCGGCGAGGCTGACGGCGATGGCGCAGGACCTGGAG
consensus AAGGTTGGGGCGCCGCGGAGGTGGCGGCGAGGCTGACGGCGATGGCGCAGGACCTGGAG

AK107637.1 CTTGCCAGCGCACGGCGCTCGGCGTCTCGGCGCGGCCACGGAGCCGGAGCTGGACCAG
NM_001057674 CTTGCCAGCGCACGGCGCTCGGCGTCTCGGCGCGGCCACGGAGCCGGAGCTGGACCAG
D16507.1 CTTGCCAGCGCACGGCGCTCGGCGTCTCGGCGCGGCCACGGAGCCGGAGCTGGACCAG
consensus CTTGCCAGCGCACGGCGCTCGGCGTCTCGGCGCGGCCACGGAGCCGGAGCTGGACCAG

AK107637.1 TTCATGGAGGCGTACCACGAGATGCTGGTGAAGTACAGGGAGGAGCTGACGAGGCCGCTG
NM_001057674 TTCATGGAGGCGTACCACGAGATGCTGGTGAAGTACAGGGAGGAGCTGACGAGGCCGCTG
D16507.1 TTCATGGAGGCGTACCACGAGATGCTGGTGAAGTACAGGGAGGAGCTGACGAGGCCGCTG
consensus TTCATGGAGGCGTACCACGAGATGCTGGTGAAGTACAGGGAGGAGCTGACGAGGCCGCTG

AK107637.1 CAGGAGGCCATGGAGTTCTTGAAGGAGGTGGAGACGCAGCTCAACACGCTCTCCATCTCC
NM_001057674 CAGGAGGCCATGGAGTTCTTGAAGGAGGTGGAGACGCAGCTCAACACGCTCTCCATCTCC
D16507.1 CAGGAGGCCATGGAGTTCTTGAAGGAGGTGGAGACGCAGCTCAACACGCTCTCCATCTCC
consensus CAGGAGGCCATGGAGTTCTTGAAGGAGGTGGAGACGCAGCTCAACACGCTCTCCATCTCC

AK107637.1 GGCAATCGCTTCGCAATATCCTTTCTCCGGCTCTTCTGAGGAGGACCAAGAAGGTAGC
NM_001057674 GGCAATCGCTTCGCAATATCCTTTCTCCGGCTCTTCTGAGGAGGACCAAGAAGGTAGC
D16507.1 GGCAATCGCTTCGCAATATCCTTTCTCCGGCTCTTCTGAGGAGGACCAAGAAGGTAGC
consensus GGCAATCGCTTCGCAATATCCTTTCTCCGGCTCTTCTGAGGAGGACCAAGAAGGTAGC

```

Figure A.12: Alignment of the nucleotide sequences of rice OSH(knox-) (AK107637.1, NM_001057674) and OSH full length mRNA (D16507.1).


```

KNAT2      TYI DCGKV GAPMEI ACI LEEI QRENHVKRDVAP- LSCFGADPELDEFMET
KNAT1      AYLDCKKI GAPPDVVDRI TAARQDFEARCCRSTPSVSASSRDPELDGFMEA
consensus  -YI DCGKv GAP- ei - - - | - - - - - e - - - - - - - - - - - P s | S - - - - DPELD- FME-

```

```

KNAT6      AYI DCGKV GAPPEI ACL LEEI QRES DVYKCEVVPS- SCFGADPELDEFMET
KNAT1      AYLDCKKI GAPPDVVDRI TAARQDFEARCCRSTPSVSASSRDPELDGFMEA
consensus  AYI DCGKv GAPPei - - - | - - - - - e - d - - - C - - - P S v S - - - - DPELD- FME-

```

Figure A.13 Amino acid alignment of KNAT1 with KNAT2 and KNAT6 in the region of epitope binding.

Like KNAT1, KNAT2 and KNAT6 belong to class I KNOX family of TALE HD proteins. KNAT1 shows an overall identity and similarity of 44% and 66% respectively with KNAT2 and 35% and 58% respectively with KNAT6 . At the N terminal end of this region KNAT1 shows an identity and similarity of 71% and 100% respectively with KNAT2 and 80% and 100% respectively with KNAT6; and at the C terminal end of this region KNAT1 shows an identity and similarity of 88% and 100% respectively with both KNAT2 and KNATA6. The antibody (sc-19215) raised against 20 aa in this region of KNAT1 detected additional band at about 35kD in the Western blotting, which corresponds to the size of KNAT2 or KNAT6 proteins.

```

KL 40 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K24 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KL 1 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S16 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K20 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KL 34 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN5 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KL 22 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S26 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S23 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN9 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN25 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KL 37 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S14 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN18 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K32 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K26 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KL 41 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN26 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KN19 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S30 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KA4 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S20 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KA13 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KA19 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KA10 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K88 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
K43 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
S18 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KI 11 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KI 1 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KF 10 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KI 14 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
KI 4 ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG
consensus ACATGGCAGAATATGGAGATGCTGGCGTCAGCTCGATGATGGGGTTGATAACCCACATG

KL 40 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K24 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KL 1 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S16 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K20 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KL 34 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN5 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KL 22 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S26 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S23 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN9 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN25 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KL 37 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S14 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN18 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K32 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K26 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KL 41 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN26 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KN19 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S30 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KA4 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S20 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KA13 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KA19 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KA10 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K88 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
K43 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
S18 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KI 11 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KI 1 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KF 10 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KI 14 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
KI 4 AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC
consensus AGCCTGAGCAAAGCGAAAAATATTATGTCCACGAGGATCCCAAGCTCGTTTTCATCATTCC

```

Figure A.14 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles. 10 out of 29 PtKN1(HD+) sequences have G instead of C at position 118 downstream of the start codon.

KL 40	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K24	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 1	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S16	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K20	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 34	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN5	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 22	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S26	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S23	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN9	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN25	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 37	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S14	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN18	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K32	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K26	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 41	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN26	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KN19	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S30	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KA4	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S20	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KA13	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KA19	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KA10	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K88	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
K43	ATTCGTC	CT	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
S18	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 11	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 1	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KF10	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 14	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 4	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
consensus	ATTCGTC	CA	GCCGTT	GGT	TAA	GACAAA	AGCT	TGCTC	T	GGTCT	TCGTAAT	CAGTT	TGAAT
KL 40	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K24	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 1	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S16	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K20	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 34	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN5	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 22	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S26	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S23	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN9	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN25	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 37	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S14	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN18	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K32	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K26	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 41	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN26	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KN19	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S30	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KA4	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S20	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KA13	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KA19	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KA10	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K88	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
K43	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
S18	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 11	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 1	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KF10	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 14	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
KL 4	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC
consensus	TTTCA	CAGGGA	ACA	ACCCGG	AAA	TTGCT	ACACAGAT	CAGT	CCTCAA	AATAT	TC	CCGCT	AAGCC

Figure A.15 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles. 10 out of 29 PtKN1(HD+) sequences T instead of A at position 310 downstream of the start codon.

KL 40	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K24	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 1	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S16	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K20	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 34	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN5	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 22	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S26	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S23	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN9	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN25	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 37	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S14	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN18	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K32	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K26	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 41	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN26	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KN19	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S30	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KA4	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S20	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KA13	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KA19	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KA10	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K88	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
K43	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
S18	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KI 11	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KI 1	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KF10	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KI 14	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KI 4	TAGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
consensus	TaGCGCACCCACAGTATCCGAGCTT GTT GGGAGCTTACATCGACTGTCAAAA GATTGGGG
KL 40	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K24	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KL 1	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S16	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K20	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KL 34	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN5	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KL 22	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S26	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S23	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN9	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN25	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KL 37	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S14	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN18	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K32	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K26	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KL 41	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN26	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KN19	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S30	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KA4	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S20	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KA13	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KA19	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KA10	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K88	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
K43	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
S18	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KI 11	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KI 1	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KF10	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KI 14	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
KI 4	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC
consensus	CTCCTCCAGAAGCTGTGGCCCGTTTGGATGCTCTAACTCGCGAACA CCAAGA CCCACAAC

Figure A.16 Alignment of the nucleotide sequences of PtKN1 cDNAs showing alleles. 13 out of 29 PtKN1(HD+) sequences T instead of A at position 579 downstream of the start codon.

APPENDIX B

GENOMIC SEQUENCES

>5` Contig

GAGAAGTACAAGGGTCTCGAAATTTTGTAGGTCGCCTGGCCAAAGAGCTGATGTCTAGGATTTTTTTTACC
CAAATGCTGGCTGAATGCTGTGAAATAAATTAAGGTTTTGAAGAAAGTGTTTCATGAATATGGAGCGTCT
GAATACAGCAGCTGCACAAGCCTCGTCTTTCGCTTTATGGAGTTAACATGGCAGAATATGGAGATGCTGGCG
TCAGCTCGATGATGGGGTTGATAACCCAACATGAGCCTGAGCAAAGCGAAAATATTATGTCCACGAGGATC
CCAAGCTCGTTTTTCATCATTCCATGGCCATGCCGATTGCCCTTCTCTCAGCAGCAATGTTTTCAGGGTTCTCA
AGGAGATCATAAGCTCAATCCACAGCCTGGGATGAACCAGCAGCTAGCCTCTGAGCAGTCTATCATGTCCAG
ATTCTGCTCATGCCGTTGGTTAAGACAAAAGCTTGGCTCTGGTCTTCGTAATCAGTTTTGAATTTACAGGGAA
CAACCCGGAAATTGCTACACAGATCAGTCCTCAAATATTCCGCTAAGCCCCATAGTCACATCGTTAGCCTC
GCAGGCTCGAGGAGAAGCGCGGATGATACCGTCCTTGGATGCCAACAGTGCTCATTTCATGTGGATAACG
AGGAGCATGCAATAAAATCGAAAATCTTAGCGCACCCACAGTATCCGAGCTTGGTGGGAGCTTACATCGAC
TGTCAAAAGGTAAAAGCAACTCGCATTAAACAATCGAAACTAGTCGTACGTGTTGCACTGTTTGTGAAGTG
TCAAACCCCTGGGATGTGTGCAAGTATGGGATCCGAATCTAGATTTACTTCTTGTGTTTGCAGATGATGCTT
GTATTGCTGTTTTCTTTTGTGTCAATTTTTAATGCCTTTGTCTGGAAAGTAACGCTAGTATCTACCATCAC
GGTTCCTGCCTGTCCAACCTTCCGGAAAAATTAGGGTTTTCTACTAAGAAAAGCTATTGTATGCAGGGAACA
TTTTAGGATTTTTTTCATTTGCATAAATGTATGGGCATTTTTCTCTGTGAAATTTTTCTGTCACTAGGTTGTT
CCGACTGCTGCCATGAAATGAATTTTAACCTTTTGGCCGCACCTGCGACTCTTGTTTTTCTTGTGGATGC
AGATTGGGGCTCTTCCAGAAGCTGTGGCCCGTTTTGGATGCTCTAACTCGCGAACACCAAGACCCACAACGT
CGGACTGTGAGCATCGGAATGGACCCAGAACTCGATCAATTTATGGTAAAACTTTTCTATCTTTTTTCAT
TTCTCTTCTTACTACTGATGCTTAAATAATTGATCAGACACAAGAGTTTTCCAAACACATTACTTGTGACT
ATTCTCTTAGTCGAAAATTGTAAAGAGGTCAATAACTACCCTGCCTGATTCTTTGAAAAAAGGTCAGAT
GGATTTCGCTTCTAAAATTACTGAAAACAAGACTTGGTCAAAGGATTTCGCTTCTGAAAATACTGGAAAGCT
AAAAAAGCACTCATGATGATGCATTCTTTCTTTACCAGGAGGCATACTGCGAGATATTGACTAAGTACCA
TGAAGAGCTTGCCAAACCTTTCAAAGAAGCAATGCTGTTTTTGAAGAAGATTGAAACCCAGTTCAACTCCC
TGGGAAAAGGAACGATAAGAATTTCTTCTCCAGGTAATACTCGCTTGCTTCATAAACATTTTTTACTTCTT
TTTTCTCTATTGCTCCAATTAATTCTGCTTCATATCAACCCACTATGTAAATTCAGTTATTTCGGGAGA
TAAAACCTTGAAGGAAATCAATGGGCTTCGAGAAAATCCAGTGAAAACCGAACTAGCTGTAAATTGCGTTAA
GTACAAGTCTAGACTGCCAAATGTAGATTATTAATGTTGAGGACAAAAATCTCAGGGCCATGGACGATAT
CGAGGGATGGAATTTAAATGAAATTTTTGTTCTTATTATTTTTGGTGTCTATATGCCACTATAATGGCA
AGACCTCAAATCATTATGGATATCATCAAGCCATGAAAGTCTTAGGGGGGTTTACAGAAAACCGCTGGGGG
GTTACTGTTACTATTCACTACTTTAAATATAAACCCAGTACTAAATATCCTGGTCAAGGGGCTTACA
GGAAACCGCTGGGAGGTTAGTGTTCGTATTCATCGCTATTTTTAAATATAAACACAGTACTAAATATTCAG
TCCTACATTAAGCATTATTTTGGTGTCAATTAATGGAGATATAGTTGATAGCTTTCATTAGTGAGTGGC
ACTTATGGTTGTAGTACGTACAGCTAGTTGTAGACTGCTTTATATAGTCTAAGATGACATTCTAGTCTTAG
GCATATGACTTGTGTAATGCCAGACTTATTATACCAGAAATGAAAATTTATATGAAGAATACTAATGCG
ACCCTTAAAAGGCTACGGTATACCGTAGATTTATAATATTACATATTATACAATGAGAAGGAAGATTGCG
CCTTACATTTGTTATATAATGCAAGACCAAAGGACCCAAAATACAGTGATAGGAAAGCTTGGAAACGTTTCA
TTTTCTTGAGATGGATTTATCCGAGGAAGATATATAACGCCCAAGTTAGCACCCCTAAACATCATCATGCT
AAAATGAAAGATTAACGATCCTAAAAGCAAGCAAATACCGTAAGTCTGTACATGAAATCTTCCCATTCACT
TTTTATCAAATTTTCATGCTTCTACACATTTATTTGCATCCATTTCTCATGGCGCCACAATCAAGGAAAAAT
AAGTGAAACAATTCTATCAAGACCAAACACTTTACTTCTGCGTAATTATATTCAATGAGTTTTGTGGCATA
GTTATTTTTGAACCATTAATGTGTTATACTAATTAACAACCTATATAATCAACCAGAAGTGATATTTGCC
ATCTCAACGAAAAATTGAACAAAAAAATTTGAAGGGATTCAATTTATATTGGGCCATTAGTATGTGCATATA
AAGATTTTTATGTGAATTTTTATATATATGACAAATAAAGCATTGTATTGGTGTGTTTTTTTTCTTGAGCTTTA
GGGTATGAGGCCTTTTGCAGAGAAATATATCCATGAATGCTCATTGGGGTTGCAAAGATAAGTGGTTCTTA
ATAGTGTTCATTGTGATTTATTAATCCTTAATTTTTTATTGTATTCTATGAGAATTAATAATGTTTTCGA
TCTTACCTTAGGTCCAACGCTTGTGTAGCAGTCCAGCCTTGCTGATCCAGATGAGCTCCTCCTTCACTTT
AAGACCTGCATACACTAACAAAGCAAACGATTACATAACATCTGAATATTCATCATAAAACAAAGCATATT
CATCACATAGACCAAGGAGTACATGAAATAATTCATTAACAACAAATCATTGAGAAGAATGGACTTATTTG
TATTAATAATCAAAAAGGAACCATACAGCCTAAGCCTTGGTGTCTGTGCAAGCTCCTCCCAGGGTCCAAC

AAGAGATTTCTCTACTTTGACAAGCTCACCCACTCCCAATATGGCAAGAAGGAAGGATAGTGATCTCTCC
TCACAAGTCAACAATGGAAGTCCTAATTCAGCAGCATTACCCCTTTAGAATGTTTCAGCTTTCAACCTCCC
CTCCTTTGGCCCAATGCCTGTATTTATAGTATTCTAGAGGAGGTTATTTCAGTTCCAACATCAATGGGATAC
TCCCTCTTTGGATGGGGTAGCTCCTCTCTTGATGGGATAAGCTCCTATATCATAGGGATGCATGAACCAT
CTGGAGGCGCTCAATACACGTCAAAAACCTCTTCTAGAAGGCTCTAGAAGTGACCACAAGTAACCTTTTTAAA
GATGAAACTTCGCGCCCTTTTGACCTTTACAGGGGTATGCACCCATCCTTTGGCTTACAGAGCGTACAAGT
ACGCAGGTACAGTGTGTTGGGGTGCGTACACGTAGGCAAGGTAGAGAGAATCTGCGTACCCAGAGATAAA
CATGTTTTTTAAGTGACTTTCTCTCTTTTTGAAGTTTTTACATGAAGAGCAGAGGTATGAGGCTTGACAAG
GAGGATATTATGACTATTGATGGAGCCATATTAAGTGCAGATAGTCATGGTGACAAATTTTCAGTGTAACAG
ATTGCCCCCTAAAGTATGAAGCCTGTAATAGATAAGGGTGTGTATCATATAGGAAAGGGTGAAGTACTTT
ATCATTTCGAGGATCATCTTGTGGAGAGCCGGTCAGAACCTGCAATTTCCAACACAAAATACATGTATAA
GATTACATCATCACTGAGGTAGGCATTTCCATACTCAGATCAAAAACCTTTAACTATTGGACTAACTCTCT
GGTACAAGAATGATTTTTATAATACATCATAATCATAGACTCGTGTAGACACAGTATGGGCGCAGCCACTCTG
TTTTATGATTGGTCTACAACATAATGAATACAAAACCTCTCGCAGCCTGGATGGCAGAAAACCTAGATCCAGGG
TCTTCATTTCTGTTCCCCGAATTTGGTACGATTCTAACCATGTGAATAAATGAGTCGATGTGCAACATGGTA
GAAATCACTTGATTACCAGTCAATGATGTAACTGTTAACAAGTGTGGGAGAGACCACCAGTAATGTCAAT
GGTTCTCAAAGTACCCAGCACAATCCATGAGATGTGAATGAACGAATCGGCATCGAACATTTTCATGGAATC
CTTCGATCGCAAGTCAATGATGAAAACCTGACGTGCAAGGCAATAAACCTGCCATTGATGCATGCAGACTCC
AAGGTACTGACGATTCCCAAGTCAACTTTCTAGATAATGTGCTGATGTCGATAGAAAGCGACTCGTCGCAG
GCAAGCCCACAACGCGAAATCATAGTTCAAGTTGAGGAAATGTTGCTGTAAGGACTGACTCAATGATGGTA
TGATGAATGGTAGGTATTTTTGCAATTTGAATGTTTTGTTTCAGATCTCTTGAACACATAACGACACAACAGA
TGGTTTTGGTGATAAATGACTCAATGTCGAATCCCATAACCCGGCTAAAAATGTCAAGTGATTTTTGGTCAAC
TGTGAGGTAAAGGATTTCCCAACCCAGGTGAATGAACAAGCGTCTGTTTCAAGAGCGGTAGTAAGGTTTTT
CTTGAGACAAAAGATATCAAATAGGCTCATTCTTCTCATTTTTTCCCCACCATTTCTGGCAAGTACCTCT
TTATTTTTTACAAGCTTGAGGAGGATAACATATGCTCGTTTGTAAAGCTCATCCTTCGGATGAGTCCAACCT
AGTTTTTGTATGCGAACCCTAAGGAGGAGCGGTCCGTTTCTGTTTGTAAATAGAACAGGTCTTTACTGTGAA
ATTTTAAACGAGTCCAGTATAGGAGAAAACCTGACGCTTTTCATATTTAGAGAGTCTGAAAAACTTAAGAAGA
CTCCAGATTTGAACTTTTTATAAAGTTGTGTCTCCTAAGCCCGAAAAAGTATAAAAATAATACCTAGGAGAA
CCGCTTGATCTGCGACCAAGGAGTACTTTTATAAATAGTGTCTTTGACCTAAAAAGTAAAAAGCGATCAA
GGACACGTCATTTTATTTCTAATATCGAATTTGAAACGACAAAAACCCGGCAATTTTTCTTGTTTTTTAC
GGTCAGAAAAGGGCTATAAGGGGACCGTCTTTTAATTCCGTTTTTAACAAGTGTAAGAAGAACGCTTGGCAA
TATTTTCTTATTTTGGGAGTCAGAAAAGGGCTTAAAAGATGACTTTGATATTTTGGTTTTGTGAAAAATGAC

>3` Contig

TGATGCATTGTAGCCCCGAGTCTTGAAAAATAAATTTAAAATCCCTCCGGCCATCCTCCGAGCTAAGTAATT
ACAGCATGGTGTAGGCCAGTAAAATATGTGATTTTGATTCAACATGGTTCATGTGCATCGAACTTTAGCTT
TGGCATGAGATTTAAAATAATAAAAATTGATAATGATTGCAGAGGTGGAATAACAGGAGAGAATAGATTCAAG
ATTGGCAATGGAGGCAACAGCGCTTCTCATCTTTATTAACAGTTGAATAAACTACATACGAGACAAGTAA
ACAAACCGCAATAAATCCCAACTTAAACTCAACTAATTTAACTAAATCACAACCTGACTATGGCACCTAGCA
CCAAACAGACACTTAAATTAATGTTCTCAACAAAAATATAGTTGTGGATGAAGCATTGCAACTGTTGCTG
AATTTCTTCTTCATCTGTTTTACAGATGGTTTTAAATTTTTCATCCACTGTTCCATATCACCTATAACCAGT
GGGGACTGTTATAGGCCTAATTTAGTACAAGAAGTGGGACCTATTTTAGTAACCTTTGGCTTAGTACGTTTG
TGTCCCTGTGATTATTTCCAGCTTATAGCCTTAATTTAGTACAAGAAGTGGAGGCAATTTTATTAAATTT
AGCTTAGTTTCGTTTGTTCCTGCGTATTCTTTTCAACAAAGAGATGTCTTTAATTTATTACTACTTTT
TCGCAGCTGACGATGACGAAAAGACGGAGGGAGGTGGATCTTCGGAGGAGGTGAGAACGGCAGTGGTGGT
GAAACGGATTTTCTGGAAGTGGATCACCATGCAGTAGAGGATAGGGAGTTAAAAATCATCTCCTTCGCAA
ATACAGTGGATATTTGAGTAGTCTAAAACAGGAATTCATGAAGAAGAAAAAGAAAGGAAAGCTCCCTAAAG
ATGCACGGCAAAAAGTTACTCGATTGGTGGAGTCTGCACGACAAGTGGCCTTATCCTTCGGTATGATTTCTC
ACTGTAAAATTTGTTTCGTTTTGTCTTTGAGAAGTGCCCCACTTTATGATATTTGTAACATAAAGGCAAA
ACCAGTATTTTACTCATGCAATGAGTTAATGTCTTCTTTTGCATTGCAGGAGACGGAGAAAATAGCTTTG
GCTGAATGCACGGGGTTGGATCAAAAACAAATAAATAATTGGTTTTATAAACCAAAGAAAACGCCACTGGAA
GCCTTCTGAAGATATGCACCTCATGGTAATGAACAGTCACAGTCCTCACAGTGTCTGCTTGTATGTTGAGA
GACATATGATGACTGAAGGGTATCTTTAGATTGCTAGATTGCTAGAAAAGAACCTTCGGCTGAAAACAGCAC
ACAATGCTATTGCTTTTTGTTGTATTTAATTGGCATGGCTTTCAATTTTCTTTCCCTTTTTTATTGGTAAGT
AAAAACTTCAAATTTATATAAAAAGGAAAGATAAAAAATAGTAATGACCCATATGCAAAGAAAAAGGTTTA
AAAGAGTGGGGAGTTACACACAGCCATAAAATGCCAGATATCGAGCAGGTACAATTCGTAAACTAAAAGGA
TAGGAGAAGAAAACAAATTTAAAAGGAAATAAAGTATAACACTTCGGCACTAGATCAGGTAGTCTCAAT

GCCGAGAGCAGTAAAACCCCTAACGGCCAAGCTAGCGTTAAACTAGTAAAACCCCTAACGGCCAAGCTAGCGT
TAAACTGAACGAGCTTAAACACAGAGGCATTGTGTCAACCGCTTCTGGGCCCTGGCACTCTCTCTCCCTCAG
GAACGATATTGGAAGAAGCTAAGTTAAAGTTGTATGCACACTTCGTAATTTTCTGACTTTTGGGAAGTAAGT
TAATCTGGAATTTTCCATTTCGAGTGGTTCCTTTGTTACTAATCTCCATGGACTTTTAATCTTGATGGATTT
ATATTTTTCTTTTATATTATCGTCTC

>Promoter

AAAAACAATTTTTATACTGCTGCTTTGAAAACCTGAAAATGGATTTTGAATGCTCTGCCCCTCACTTCACT
GCTCCAGGCAACATGAATGGGCAGGTGGACGGAGGTGTGGCCTCGGGGTGAGCCGTCATTGTTCATCGTTCGG
CGCCCGATATGGCATAGCACCCGCCCCCTGCCGCTGGCCGCGCTGTTAACCGTCATTCTCATCGTTCGGCGCC
CGACATGGCATAGCACCGTCCCCCGCGGCGATCCACGTACGTGTAGCTGATCGTTACAATCTGCTCTCCG
CACGTGACTAACAGAGGCGATTTATTACTGGTTGAGGTGTGGGGCAAGGGTCAGGGGTGGAGTATGTCATT
CACGAGCTCGTAGTTTCGAGAAGGTAGAGAGAAAAGTATTTGGTTTTGTAAGCGGTGGCGTGGGGAGCGAGGG
GCTGTGCGTTTTATTTCTGTTGAGGCGACACCGTCTTCCAAAAAAGAAGCTCGAAAATCAAGACAGAGTGGGA
GGATGTTGAGGTGGAGTTGCGACTCTCAAGAACGAAGGGGGAAAGAGGTTGGGCAATATTTATGAATATAGT
AGTACAGAGGAGTCAGGTTAGGGAGATCCTGAGGGATTCCCTTGAATGCAAACAGTACGTTTTTGGTACTT
TGAACAGATTAAACTATTCGCTGAGCCTATCGACCTTTGTTGTGTCAGCGTCGCCGGGAATGACCAGAACA
GCCCAGCCCCCGGGGATCGGGGATTTGGGAAAATGGAGAGCTGGCGGCGAAAGAAGAATAAGGGATTTTA
ACATTTACTAAGTTAAAGAAGCAGAGATTTAAGGTTTAAGGTTTACTAAGTTATGGAAGAAGAAGGAGAA
GTGATTTCAAAGTTTATAGGTTAAAGAAGAAGATATTTAATTTTTATCAAGTAAAGGAAAATATATTTTA
ATGTTTACTAAGTTAAAAATAATTTATAAGAAAATTTGTTTTTATCTGATTTTATATTCAACTTGTTTTC
TTAACCAACGTCACACTAACTATTTGGCTATCTAATTTTTCCATTCAACCCTTTTCTTATACAACCTGGAC
AATAAATTTAACACTTTTTCTTATACAACCCGACAAAAAGCTTTTCTAATTTCTCAATATTTCTCTCACCAC
ATAAATTACCCAAATTACCATGAACTAGTTATCAGGATAACTGTCATACTTTTGAAGCCTCACATCCTAA
AAATGAAATGGGCGTTGGGCATATACTTTGCTTTGCTGTTTTCTGGGTGCGAGACGCGGGTGCGGGTCCT
GCAAGACAAGATGCTGCAAGTGGTGGTGTGTGGGTTTTGTTGTACTGGGCCAGCATGATTTTGTTCGG
AGGATTGCACTGCAGCTGAACATTAAGACAGGCCACATGATTTTGAAGGAAAGGAAAAGACATAACGGC
ATCTCTGTTGGGATCCAGAATTGAGAAAGGGACGCTGGTCGGTTTTCCACAATGAATGAATGAAGGCCACCA
GTGCATATGAGACTTTCTCTTCTATTTGTAGAGGATAAGTCGGGGATTGAAATATATGTTTCGCCAACAGAC
TGACAAGAGGAACTTAGAGCCTATGGACTGCCGATGAAATGGGAGAAGTACAAGGGTCTCGAAATTTTGT
AGGTCGCTGGCCAAAGAGCTGATGTCTAGGATTTTTTTTACCCAAATGCTGGCTGAATGCTGTGAAATAA
ATTAAGGTTTTGAAGAAAGTGTTCATGAATATGGAGCGTCTGAATACAGCAGCTGCACAAGCCTCGTCT
TCGC

APPENDIX C

PROTOCOLS

RNA isolation by TRI reagent

(TR-118, Molecular Research Center, Inc.)

Embryos from different stages of pine embryo development, early stages 1-4, middle stages 5-8, and late stages 9.2-9.3, 9.4-9.5, 9.6-9.7 and 9.8-9.9 were pooled separately and RNA was isolated using TRI-reagent protocol.

- 1) Powder the tissue and add 1000 TRI Reagent (TR-118, Molecular Research Center, Inc.).
- 2) Store at room temperature for 5 minutes.
- 3) Add 100 μ l BCP (1-bromo-3-chloropropane, ACROS) and vigorously shake for 15 seconds.
- 4) Store at room temperature for 15 minutes.
- 5) Centrifuge for 15 minutes at 14,000 rpm at 4°C in the Beckman F2402 rotor.
- 6) Collected the aqueous phase for RNA isolation into a DEPC treated 1.5 ml tube.
- 7) Add 250 μ l of isopropanol and 250 μ l high salt precipitation buffer (0.8 M sodium citrate and 1.2 M NaCl).
- 8) Mix by inverting and store for 5-10 minutes at room temperature.
- 9) Centrifuge for 10 minutes at 14000 rpm at 4°C.
- 10) Wash the pellet with 1.0 ml 75% ethanol.
- 11) Centrifuge for 5 minutes and air-dry in the fume hood about 5 minutes.
- 12) Resuspend the pellet in 30ul RNase/DNase-free water and incubated for 10 minutes at 55°C.

SMART RACE

(Rapid Amplification of cDNA Ends - Clontech, MountainView, CA)

Once RNA was isolated from different stages of pine embryos, RACE ready cDNA was prepared using SMART RACE amplification kit.

- 1) 500ng of RNA was used for each pooled stages.
- 2) For 3` RACE ready cDNA, add 1ul of 3` CDS primer to 1-3ul of RNA sample and make up the volume upto 5ul with water.
- 3) For 5` RACE ready cDNA, add 1ul of 5` CDS primer and 1ul of SMART II A oligo to 1-3ul of RNA sample and make up the volume upto 5ul with water.
- 4) Spin briefly in a microcentrifuge to collect the volume at the bottom of the tube.
- 5) Incubate the tubes at 70°C for 2 minutes and then cool them on ice for 2 minutes.
- 6) Add 2ul of 5X first-strand buffer, 1ul of DTT, 1ul of dNTP and 1ul of powerscript reverse transcriptase.
- 7) Spin briefly in a microcentrifuge to collect the volume at the bottom of the tube.

- 8) Incubate the tubes at 42°C for 1.5 hours.
- 9) Add 100ul of Tricine-EDTA buffer and heat the tubes at 72°C for 7 minutes.
- 10) Store the samples in -20°C.

DNA and gel purification

Once the PCR products are run on gel to confirm the size of the band of interest, either the band is cut out of the gel or remaining PCR products are purified using either of the following kit for cloning.

(QIAquick kit from Qiagen - Valencia, CA)

- 1) Add 3 volumes of the buffer QG to 1 volume of gel and incubate at 50°C for 10 minutes or until gel melts or add 5 volumes of buffer PBI to 1 volume of the PCR product.
- 2) Transfer the mix to QIAquick column and centrifuge at 14000 rpm for 1 minute.
- 3) Discard the flow-through collected in the collection tube.
- 4) Add 750ul of buffer PE and centrifuge at 14000 rpm for 1 minute.
- 5) Discard the flow-through and centrifuge at 14000 rpm for 1 minute to remove residual ethanol.
- 6) Transfer the column to clean 1.5ml microcentrifuge tube.
- 7) Add 30ul of buffer EB and let the column stand for 1 minute.
- 8) Centrifuge at 14000 rpm for 1 minute and the volume collected is the purified DNA.

(Wizard SV gel and PCR clean up kit from Promega - Madison, WI)

- 1) Add 1 volume of the DNA binding solution to 1 volume of PCR product or to 1 volume of gel and incubate the gel containing mix at 50°C for 10 minutes or until gel melts.
- 2) Transfer the mix to SVmini column and centrifuge at 14000 rpm for 1 minute.
- 3) Discard the flow-through collected in the collection tube.
- 4) Add 700ul of column wash buffer and centrifuge at 14000 rpm for 1 minute.
- 5) Discard the flow-through.
- 6) Add 500ul of column wash buffer and centrifuge at 14000 rpm for 5 minutes.
- 7) Discard the flow-through and centrifuge at 1400 rpm for an additional minute to remove residual ethanol.
- 8) Transfer the column to clean 1.5ml microcentrifuge tube.
- 9) Add 30ul of nuclease free water and let the column stand for 1 minute.
- 10) Centrifuge at 14000 rpm for 1 minute and the volume collected is the purified DNA.

Cloning

(pGEM-T or p-GEM T Easy vector system from Promega - Madison, WI)

Ligation

- 1) Once the DNA is purified it is cloned into pGEM-T Easy vector.
- 2) Add the following into a 0.5ml microcentrifuge tube - 5ul Rapid Ligation Buffer + 1ul of pGEM-T or T Easy vector + 3ul of purified DNA product + 1ul of T4 DNA Ligase.
- 3) Incubate overnight at 4 °C.

Transformation

- 1) Take 2ul of ligation mix into a 1.5ml microcentrifuge tube and add 30ul of JM109 competent cells to it.
- 2) Incubate on ice for 20 minutes.
- 3) Heat-shock the cells at 42 °C for 45 seconds.
- 4) Immediately put them back on ice and incubate for 2 minutes.
- 5) Add 950ul of SOC medium to the tube and incubate at 37 °C for 1.5 hour while shaking.
- 6) Centrifuge at 2000 rpm of 10 minutes.
- 7) Discard the supernatant and add 200 ul of SOC to the pellet.
- 8) Spread 100ul of cells on 2 LB Agar plates containing Ampicillin, IPTG and X-Gal each at a final concentration of 100µg/ml.
- 9) Incubate the plates overnight at 37°C. The white colonies that appear are the ones with insert and blue colonies are empty vector.

Colony PCR

(GoTaq Green Master Mix from Promega - Madison, WI)

- 1) Prepare master mix of 50 ul as follows – 25ul of the green mastermix + 23ul of H₂O + 1ul T7 primer + 1ul SP6 primer.
- 2) Aliquot 10ul of the mix into PCR tubes for each colony to be checked.
- 3) Touch a white colony with sterilized toothpick then touch onto a LB-Agar plate and then dip it into the PCR mix.
- 4) Repeat the same for as many colonies to be checked.
- 5) Run the PCR in Eppendorf thermocycler and run on gel when ready. Grow the colonies overnight at 37 °C.

Plasmid preparation and purification

(Wizard SV minprep kit from Promega - Madison, WI)

Prepare bacterial culture

- 1) Take 2ml of Terrific Broth into a 15ml falcon tube and inoculate colony of interest in sterile conditions.
- 2) Incubate at 37°C for 16 hours.

- 3) Transfer 500ul of the culture into a 1.5ml microcentrifuge tube and add 500ul of LB-Glycerol and flash freeze in liquid nitrogen. Store the glycerol stock at -80 °C.
- 4) Use the remaining culture to purify plasmid.

Plasmid purification

- 1) Centrifuge the culture at 2000 rpm for 5 minutes.
- 2) Discard the supernatant and add 250ul of resuspension solution.
- 3) Vortex carefully and transfer into a 1.5ml microcentrifuge tube.
- 4) Add 250ul of cell lysis buffer. Mix by inverting 4 times and incubate at room temperature for 1-5 minutes.
- 5) Add 10ul of alkaline protease solution. Mix by inverting 4 times and incubate at room temperature for 5 minutes.
- 6) Add 350ul of neutralization buffer. Mix by inverting 4 times and centrifuge at 14000 rpm for 10 minutes.
- 7) Transfer the supernatant into a SVmini column placed over a collection tube.
- 8) Centrifuge at 14000 rpm for 1 minute.
- 9) Discard the flow-through. Add 750ul of column wash buffer and centrifuge at 14000 rpm for 1 minute.
- 10) Discard the flow-through. Add 250ul of column wash buffer and centrifuge at 14000 rpm for 1 minute.
- 11) Discard the flow-through and centrifuge at 14000 rpm for an additional minute to remove residual ethanol.
- 12) Transfer the column to a fresh 1.5ml microcentrifuge tube and add 30ul of nuclease free water to the column.
- 13) Let the column sit for 5 minutes at room temperature and then centrifuge at 14000 rpm for 1 minute.
- 14) The volume collected is purified plasmid. Measure absorbance at 260/280nm.

Protein isolation by TRI reagent

(TR-118, Molecular Research Center, Inc.)

Embryos from different stages of pine embryo development, early stages 1-4, middle stages 5-8, and late stages 9.2-9.3, 9.4-9.5, 9.6-9.7 and 9.8-9.9 were pooled separately and protein was isolated using TRI-reagent protocol.

- 1) Tissue was powdered followed by addition of 1000 µl TRI Reagent (TR-118, Molecular Research Center, Inc.) then stored for 5 minutes at room temperature.
- 2) 100 µl BCP (1-bromo-3-chloropropane, ACROS) was added and vigorously shaken for 15 seconds then stored at room temperature for 15 minutes followed by centrifugation for 15 minutes at 14,000 rpm at 4° C in the Beckman F2402 rotor.
- 3) Collect the organic phase (phenol-ethanol) at the bottom collected into a microfuge tube.
- 4) Add 3 volumes of acetone to precipitate proteins. Mix by inversion for 10-15 seconds to obtain a homogeneous solution.
- 5) Store the samples at room temperature for 10 minutes.

- 6) Centrifuge at 12,000 g for 10 minutes at 4 C to precipitate the protein.
- 7) Decant the supernatant and add 0.5 ml of 0.3 M guanidine hydrochloride in 95% ethanol + 2.5 % glycerol (V:V) to the pellet.
- 8) Disperse the pellet using a pipet tip and by vortexing.
- 9) Add another 0.5 ml of the guanidine hydrochloride/ethanol/glycerol wash solution to the sample and store at room temperature for 10 min.
- 10) Centrifuge the protein at 8,000 g for 5 min and decant the wash solution.
- 11) Repeat the wash two more times in 1 ml each of the guanidine/ethanol/glycerol wash solution.
- 12) Disperse the pellet by vortexing after each wash to efficiently remove residual phenol.
- 13) Wash in 1 ml of ethanol containing 2.5 % glycerol (V:V).
- 14) Store at room temperature for 10 min.
- 15) Centrifuge the protein at 8,000 g for 5 minutes.
- 16) Decant the supernatant and air-dry the pellet for 7-10 min at room temperature.
- 17) Add 200-1000ul sample buffer sample buffer containing 10% glycerol, 2% SDS, 5% b-mercaptoethanol depending on the pellet size.
- 18) Solubilize the pellet by flicking the tube or pipetting.
- 19) Store the protein sample at -20°C

Protein Concentration Assay

Biorad Protein Assay buffer was used to measure protein concentrations.

- 1) Dilute the dye reagent to 1X concentration.
- 2) Prepare BSA (Bovine Serum Albumin) standards (0.25, 0.5, 1.0, 1.5, 2.0 mg/ml) and protein samples for reading - 20ul of each in 1ml of the diluted dye reagent. Use water for blank.
- 3) Incubate for 5 minutes at room temperature and measure absorbance (triplicate) using a spectrophotometer at visible light setting.

Western Blotting

Protein Gel

- 1) Use 4-20% precast protein gel from Biorad separate proteins by electrophoresis.
- 2) Load 15ug of protein samples to the well and load the marker
- 3) Run at 100-120 V for 1hour or more until the dye marker reached the bottom.

Protein Transfer

Mini Trans-Blot system from bio-rad was used for protein transfer.

- 1) Presoak 2 filter pads, 6 filter papers and a nitrocellulose membrane (slightly larger in size than the gel) in transfer buffer for 15-20 minutes.
- 2) Presoak the gel in transfer buffer for 15-20 minutes.
- 3) Assemble sandwich in order as indicated below on one side of the transfer cassette. Roll a test tube over after putting the filter papers, membrane and gel to remove the bubbles.

- a. Filter pad
 - b. Whatman filter paper
 - c. Nitrocellulose paper
 - d. SDS-PAGE gel
 - e. Whatman filter paper
 - f. Filter pad
- 4) Close the transfer cassette, place it in the electrode module and put the module in the buffer tank filled with pre-chilled transfer buffer.
 - 5) Place the frozen bio-ice cooling unit next to the module. Add a stir bar to keep buffer temperature even.
 - 6) Put the lid on the tank and hook up the leads (negative to negative and positive to positive) with the membrane being on positive side and gel on negative.
 - 7) Set the transfer at 20 V for overnight at 4°C.

Blotting

- 1) Block the membrane in blocking solution (5% non-fat dry milk in TBS-T) for 1 hour.
 - 2) Rinse twice and wash 1 X 15 minutes and 2X 5 minutes in TBS-T.
 - 3) Prepare 1° antibody solution – 1: 500.
 - 4) Incubate the membrane in 1° antibody solution for 1 hour or more.
 - 5) Rinse twice and wash 1 X 15 minutes and 2X 5 minutes in TBS-T.
 - 6) Prepare 2° antibody solution (1:3000 – Goat Anti-rabbit antibody).
 - 7) Incubate the membrane in 2° antibody solution for 1 hour.
 - 8) Rinse twice and wash 1 X 15 minutes and 2X 5 minutes in TBS-T.
- Note - Do all incubation and washes at room temperature.

Detection

Chemiluminescent detection

(Immun-Star Goat Anti-Rabbit (GAR)-AP Chemiluminescent Detection Kit)

Prepare detection solution – 3ml immune-star chemiluminescent substrate + 150ul immune-star enhancer.

Incubate the membrane in the detection solution in a hybridization bag for 5 minutes at 37°C.

View the membrane in luminescent image analyzer and take a picture

Image analyzer used - Fujifilm luminescent image analyzer LAS-1000 plus system

Color Substrate detection

(Immun-Blot Goat Anti-Rabbit (GAR)-AP Detection Kit)

Prepare detection solution – 10 ml 1X AP color development buffer + 100ul color reagent A which is BCIP (5-bromo-4-chloro-3-indolyl phosphate) + 100ul color reagent B which is NBT (nitroblue tetrazolium).

Incubate the membrane in detection solution until color develops.

Preparing genomic libraries for genomic walk

(Universal Genome Walker Kit - Clontech, MountainView, CA)

Check quality of Genomic DNA

- 1) Check the size of genomic DNA on a 0.6% agarose gel with EtBr. DNA should run as large bands above 50kb.
- 2) Check the purity of genomic DNA by Dra I digestion.
- 3) Add 5 μ l genomic DNA, 1.6 μ l Dra I (10 units/ μ l), 2 μ l 10X Dra I Restriction buffer and 11.4 μ l deionized H₂O to a tube and mix gently. Spin briefly to collect the volume at bottom of the tube.
- 4) Also set up a control digestion without enzyme.
- 5) Incubate at 37°C overnight.
- 6) Run 5 μ l of each reaction on a 0.6% agarose gel with EtBr. Smear indicates that DNA can be digested by restriction enzymes.

Digestion of Genomic DNA

- 1) For each library construction, set up a total of five reactions. Set up four blunt-end digestions for each restriction enzyme provided (EcoRI, DraI, PvuII, StuI). Use human genomic DNA from kit as a positive control.
- 2) Add 25 μ l genomic DNA (0.1 μ g/ μ l), 8 μ l restriction enzyme (10 units/ μ l), 10 μ l enzyme buffer (10X), 57 μ l deionized H₂O to separate tubes for each reaction and mix gently.
- 3) Incubate at 37°C for 2 hours.
- 4) Vortex the reaction mix at slow speed for 5–10 seconds and put them back to 37°C overnight (16–18 hr).
- 5) Run 5 μ l from each sample on a 0.6% agarose/EtBr gel to determine if digestion is complete.

Purification of DNA

- 1) Gel and PCR purification kit from promega was used for DNA purification.
- 2) To each reaction tube, add an equal volume DNA binding solution from the kit.
- 3) Transfer the volume into the column provided in the kit.
- 4) Centrifuge at room temperature for 1 minute at 14000 g.
- 5) Discard the flow through collected in the collecting tube.
- 6) Add 700ul of column wash solution to the column and centrifuge for 1 minute at 14000 g.
- 7) Discard the flow through and add 200ul of wash solution.
- 8) Centrifuge for 5 minutes at 14000 g.
- 9) Discard the flow through and repeat the centrifugation step for additional minute at 14000 g to remove residual ethanol.
- 10) Transfer the column on to a fresh 1.5ml tube and add nuclease-free water on top of the column.
- 11) Incubate at room temperature for 5 minutes and centrifuge at 14000 g for 1 minute.
- 12) The volume collected in the tube is purified DNA.

Ligation of Genomic DNA to GenomeWalker Adaptors

- 1) Set up five ligation reactions for genomic DNA with 4 restriction enzymes and 1 human genomic DNA control.
- 2) Transfer 4 μl of digested product from each tube to a fresh 0.5-ml tube.
- 3) Add 1.9 μl GenomeWalker Adaptor (25 μM), 1.6 μl 10X Ligation Buffer, 0.5 μl T4 DNA Ligase (6 units/ μl) to each tube.
- 4) Incubate at 16°C overnight.
- 5) Stop the reaction by incubating at 70°C for 5 min.
- 6) Add 72 μl of TE (10/1, pH 7.5). Vortex at slow speed for 10–15 sec.
- 7) Store the genomic libraries at -20°C.

REFERENCES

1. Suhultz, *Loblolly: the pine for the twenty-first century*. New forests, 1999. **17**: p. 71-88.
2. [cited 2008; Available from: <http://www.afandpa.org/>].
3. Little, E.L.J., *Atlas of United States trees*. . Conifers and important hardwoods., ed. U.S.D.A.M.P. 1146. Vol. 1. 1971: Washington, D.C.
4. Baker, J. and O. Langdon, *Pinus taeda L.* . Loblolly Pine, ed. B.H. In RM Burns. Vol. 1. 1990, Washington, D.C: Silvics of North America, Conifers, USDA For Service Agriculture Handbook 654.
5. Ragauskas, A.J., et al., *The path forward for biofuels and biomaterials*. Science, 2006. **311**(5760): p. 484-9.
6. Timmis, *Bioprocessing for tree production in the forest industry: conifer somatic embryogenesis*. Biotechnol Prog, 1998. **14**: p. 156-166.
7. Grossnickle, S. and B.C.S. Sutton, *Application of biotechnology for forest regeneration*. New Forests. Vol. 17. 1999, Netherland: Kluwer Academic Publishers. 213-226.
8. Walter, C., et al., *Review: Application of biotechnology to forestry - molecular biology of conifers* World Journal of Microbiology & Biotechnology, 1998. **14**: p. 321-330.
9. Steward, F.C., *Growth and Organized Development of Cultured Cells. III. Interpretations of the Growth from Free Cell to Carrot Plant* American Journal of Botany, 1958. **45**(10).
10. Gupta, P.K. and D.J. Durzan, *Biotechnology of Somatic Polyembryogenesis and Plantlet Regeneration in Loblolly Pine*. BioTechnology, 1987. **5** p. 147-151.
11. Pullman, G.S. and Webb. *An embryo staging system for comparison of zygotic and somatic embryo development*. in *Proceedings of the TAPPI R&D Division Biological Sciences*. 1994. Minneapolis, MN: Technical Association of the Pulp and Paper Industry Press, Atlanta, GA.
12. Fuentes, S.R.L., et al., *The effects of silver nitrate and different carbohydrate sources on somatic embryogenesis in Coffea canephora*. Plant Cell, Tissue and Organ Culture, 2000. **60**: p. 5-13.

13. Becwar, M.R. and G.S. Pullman, *Somatic embryogenesis in loblolly pine (Pinus taeda L.)*. Somatic Embryogenesis in Woody plants., ed. G.P. Jain S, Newton RJ. Vol. 3. 1995: Kluwer, The Netherlands. 287–301.
14. Pullman, G.S., et al., *Improving loblolly pine somatic embryo maturation: comparison of somatic and zygotic embryo morphology, germination, and gene expression*. Plant Cell Rep, 2003. **21**(8): p. 747-58.
15. Pullman, G.S., et al., *Somatic embryogenesis in loblolly pine (Pinus taeda L.) and Douglas fir (Pseudotsuga menziesii): Improving culture initiation with MES pH buffer, biotin, and folic acid*. Plant Cell, Tissue and Organ Culture 2005. **80**: p. 91-103.
16. Pullman, G.S., K. Namjoshi, and Y. Zhang, *Somatic embryogenesis in loblolly pine (Pinus taeda L.): improving culture initiation with abscisic acid and silver nitrate*. Plant Cell Rep, 2003. **22**(2): p. 85-95.
17. Cairney, J., et al., *Natural and somatic embryo development in loblolly pine.: Gene Expression Studies Using Differential Display and DNA Arrays*. Applied Biochemistry and Biotechnology, 1999. **77-79**(28): p. 5-17.
18. Cairney, J., et al., *Transcript profiling: a tool to assess the development conifer embryos*. In Vitro Cell Dev Biol Plant, 2000. **36**: p. 155-162.
19. Jurgens, G. and U. Mayer, *Arabidopsis*. Embryos: a color atlas of developing embryos, ed. J. Bard. 1994, London: Wolfe Publishing.
20. Goldberg, R.B., de Paiva, G., and Yadegari, R., *Plant Embryogenesis: Zygote to Seed*. Science, 1994. **266**: p. 605-614.
21. West, M.A.L. and J.J. Harada, *Embryogenesis in higher plants : an overview*. Plant Cell Rep, 1993. **5**: p. 1361-1369.
22. Grossniklaus, U., et al., *Maternal control of embryogenesis by MEDEA, a Polycomb group gene in Arabidopsis*. Science, 1998. **280**: p. 446–450.
23. Lou, M., et al., *Genes controlling fertilization-independent seed development in Arabidopsis thaliana*. Proc Natl Acad Sci USA, 1999. **96**: p. 296–301.
24. Ohad, N., et al., *A mutation that allows endosperm development without fertilization*. Proc Natl Acad Sci USA, 1996. **93**: p. 4223–4228.
25. Singh, H., *Embryology of gymnosperms*. Handbuch der Pflanzenanatomie, ed. C.Z. Zimmerman W, Ozenda P & Wulff HD. 1978, Berlin, Stuttgart: Gebruder Borntraeger. 187-241.

26. Berleth, T. and S. Chatfield, *Embryogenesis: pattern formation from a single cell*. The Arabidopsis book. American Society of Plant Biologists, ed. M.E. Somerville CR. 2002, Rockville, MD.
27. Dogra, P.D., *Seed sterility and disturbances in embryogeny in conifers with particular reference to seed testing and tree breeding in Pinaceae*. Studia Forestalia Suecica, 1967. **45**: p. 5-97.
28. Nagmani, R., et al., *Anatomical comparison of somatic and zygotic embryogeny in conifers*. Somatic Embryogenesis in Woody plants, ed. G.P. Jain SM, Newton RJ. 1995, Dordrecht: Kluwer Academic Publishers.
29. Filonova, L.H., et al., *Programmed cell death eliminates all but one embryo in a polyembryonic plant seed*. Cell Death Differ, 2002. **9**: p. 1057–1062.
30. von Arnold, S.S., I. Bozhkov, P. Dyachok, J. Filonova, L., *Developmental pathways of somatic embryogenesis*. Plant Cell, Tissue and Organ Culture 2002. **69**: p. 233-249.
31. Cairney, J.Z., L. Cowels, A. Hsiao, J. Zismann, V. Liu, J. Ouyang, S. Thibaud-Nissen, F. Hamilton, J. Childs, K. Pullman, GS. Zhang, Y. Oh, T. Buell, R., *Expressed sequence tags from loblolly pine embryos reveal similarities with angiosperm embryogenesis*. Plant Molecular Biology, 2006. **62**: p. 485-501.
32. Bertolino, E., et al., *A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif*. J Biol Chem, 1995. **270**(52): p. 31178-88.
33. Burglin, T.R., *The evolution of homeobox genes*. Biodiversity and Evolution, ed. R. Arai, Kato, M. and Doi, Y. 1995, Tokyo: The National Science Museum Foundation. 291–336.
34. Kerstetter, R., et al., *Sequence analysis and expression patterns divide the maize knotted1-like homeobox genes into two classes*. Plant Cell, 1994. **6**(12): p. 1877-87.
35. Quaedvlieg, N., et al., *The homeobox gene ATH1 of Arabidopsis is derepressed in the photomorphogenic mutants cop1 and det1*. Plant Cell, 1995. **7**(1): p. 117-29.
36. Steelman, S., Moskow, J. J., Muzynski, K., North, C., Druck, T., Montgomery, J. C., Huebner, K., Daar, I. O. and Buchberg, A. M., *Identification of a conserved family of Meis1-related homeobox genes*. Genome Res., 1997. **7**: p. 142–156.
37. Burglin, T.R. and G. Ruvkun, *New motif in PBX genes*. Nat Genet, 1992. **1**(5): p. 319-20.

38. Gomez-Skarmeta, J.L., et al., *Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes*. *Cell*, 1996. **85**(1): p. 95-105.
39. Mukherjee, K. and T.R. Burglin, *Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution*. *J Mol Evol*, 2007. **65**(2): p. 137-53.
40. Simon, R., *Functions of plant shoot meristems*. *Cell & Developmental Biology*, 2001. **12**: p. 357-362.
41. Steeves, T.A. and I.M. Sussex, *Patterns of Plant Development*, ed. Cambridge. 1989, New York Cambridge Univ. Press. 86–99.
42. Hagemann, W., *Studien zur Entwicklungsgeschichte der Angiospermenblätter*. *Bot. Jahrb. Syst. Pflanzengesch. Pflanzengeogr*, 1970. **90**: p. 297-413.
43. Kaplan, D.R., *Comparative developmental evaluation of the morphology of unifacial leaves in the monocotyledons*. *Bot. Jahrb. Syst. Pflanzengesch. Pflanzengeogr.*, 1975. **95**: p. 1-105.
44. Chuck, G., C. Lincoln, and S. Hake, *KNAT1 induces lobed leaves with ectopic meristems when overexpressed in Arabidopsis*. *Plant Cell*, 1996. **8**(8): p. 1277-89.
45. Larson, P.R., *The vascular cambium*. *Development and Structure*. 1994, Berlin: Springer-Verlag.
46. Long, J.A. and M.K. Barton, *The development of apical embryonic pattern in Arabidopsis*. *Development*, 1998. **125**(16): p. 3027-35.
47. Jackson, D., B. Veit, and S. Hake, *Expression of maize KNOTTED1 related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot*. *Development*, 1994. **120**: p. 405–413.
48. Smith, L., et al., *A dominant mutation in the maize homeobox gene, Knotted-1, causes its ectopic expression in leaf cells with altered fates*. *Development*, 1992. **116**: p. 21–30.
49. Smith, L.G., D. Jackson, and S. Hake, *The expression of knotted1 marks shoot meristem formation during maize embryogenesis*. *Developmental Genetics*, 1995. **16**: p. 344-348.
50. Hake, S., et al., *The role of knox genes in plant development*. *Annu Rev Cell Dev Biol*, 2004. **20**: p. 125-51.
51. Scofield, S. and J.A. Murray, *KNOX gene function in plant stem cell niches*. *Plant Mol Biol*, 2006. **60**(6): p. 929-46.

52. Webb, D. *Mature Embryo in Seed*. [cited 2008; Available from: http://www.botany.hawaii.edu/faculty/webb/Bot201/Angiosperm/capsella_flower.htm.
53. Kerstetter, R.A., et al., *Loss-of-function mutations in the maize homeobox gene, knotted1, are defective in shoot meristem maintenance*. *Development*, 1997. **124**(16): p. 3045-54.
54. Sinha, N.R., R.E. Williams, and S. Hake, *Overexpression of the maize homeo box gene, KNOTTED-1, causes a switch from determinate to indeterminate cell fates*. *Genes Dev*, 1993. **7**(5): p. 787-95.
55. Barton, M.K. and R.S. Poethig, *Formation of the shoot apical meristem in Arabidopsis thaliana: an analysis of development in the wild type and in the shoot meristemless mutant*. *Development*, 1993. **119**: p. 823-831.
56. Long, J.A., et al., *A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis*. *Nature*, 1996. **379**(6560): p. 66-9.
57. Clark, S.E., et al., *The CLAVATA and SHOOT MERISTEMLESS loci competitively regulate meristem activity in Arabidopsis*. *Development*, 1996. **122**(5): p. 1567-75.
58. Endrizzi, K., et al., *The SHOOT MERISTEMLESS gene is required for maintenance of undifferentiated cells in Arabidopsis shoot and floral meristems and acts at a different regulatory level than the meristem genes WUSCHEL and ZWILLE*. *Plant J*, 1996. **10**(6): p. 967-79.
59. Lenhard, M., G. Jurgens, and T. Laux, *The WUSCHEL and SHOOTMERISTEMLESS genes fulfil complementary roles in Arabidopsis shoot meristem regulation*. *Development*, 2002. **129**(13): p. 3195-206.
60. Lincoln, C., et al., *A knotted1-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants*. *Plant Cell*, 1994. **6**(12): p. 1859-76.
61. Dean, G., S. Casson, and K. Lindsey, *KNAT6 gene of Arabidopsis is expressed in roots and is required for correct lateral root formation*. *Plant Mol Biol*, 2004. **54**(1): p. 71-84.
62. Gallois, J.L., et al., *Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis*. *Development*, 2002. **129**(13): p. 3207-17.
63. Pautot, V., et al., *KNAT2: evidence for a link between knotted-like genes and carpel development*. *Plant Cell*, 2001. **13**(8): p. 1719-34.
64. Douglas, S.J., et al., *KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis*. *Plant Cell*, 2002. **14**(3): p. 547-58.

65. Venglat, S.P., et al., *The homeobox gene BREVIPEDICELLUS is a key regulator of inflorescence architecture in Arabidopsis*. Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4730-5.
66. Nishimura, A., et al., *Over-expression of tobacco knotted1-type class1 homeobox genes alters various leaf morphology*. Plant Cell Physiol, 2000. **41**(5): p. 583-90.
67. Williams-Carrier, R.E., et al., *Ectopic expression of the maize kn1 gene phenocopies the Hooded mutant of barley*. Development, 1997. **124**(19): p. 3737-45.
68. Sentoku, N., Y. Sato, and M. Matsuoka, *Overexpression of rice OSH genes induces ectopic shoots on leaf sheaths of transgenic rice plants*. Dev Biol, 2000. **220**(2): p. 358-64.
69. Chuck, G., Lincoln, C. and Hake, S., *KNAT1 induces lobed leaves with ectopic meristems when overexpressed in Arabidopsis*. Plant Cell, 1996. **8**: p. 1277-1289.
70. Sundas-Larsson, A., et al., *A homeobox gene with potential developmental control function in the meristem of the conifer Picea abies*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 15118-22.
71. Belmonte, M.F., et al., *Overexpression of HBK3, a class I KNOX homeobox gene, improves the development of Norway spruce (Picea abies) somatic embryos*. J Exp Bot, 2007. **58**(11): p. 2851-61.
72. Hjortswang, H., et al., *KNOTTED1-like homeobox genes of a gymnosperm, Norway spruce, expressed during somatic embryogenesis*. Plant Physiology and Biochemistry, 2002. **40**: p. 837-843.
73. Gehring, W.J., M. Affolter, and T. Burglin, *Homeodomain proteins*. Annu. Rev. Biochem., 1994. **63**: p. 487-526.
74. Kissinger, C.R., et al., *Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions*. Cell, 1990. **63**(3): p. 579-90.
75. Gehring, W.J., *Homeoboxes in the study of development*. Science, 1987. **236**: p. 1245-52.
76. Latchman, D.S., *Transcription factors: an overview*. Int J Biochem Cell Biol, 1997. **29**(12): p. 1305-12.
77. Affolter, M., A. Schier, and W.J. Gehring, *Homeodomain proteins and the regulation of gene expression*. Curr Opin Cell Biol, 1990. **2**(3): p. 485-95.
78. Hayashi, S. and M.P. Scott, *What determines the specificity of action of Drosophila homeodomain proteins?* Cell, 1990. **63**: p. 883-894.

79. Nourse, J., et al., *Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor.* Cell, 1990. **60**(4): p. 535-45.
80. Kamps, M.P., et al., *A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL.* Cell, 1990. **60**(4): p. 547-55.
81. Kelly, M., et al., *Four mating-type genes control sexual differentiation in the fission yeast.* EMBO J, 1988. **7**(5): p. 1537-47.
82. Bellmann, R. and W. Werr, *Zmhox1a, the product of a novel maize homeobox gene, interacts with the Shrunken 26 bp feedback control element.* EMBO J, 1992. **11**(9): p. 3367-74.
83. Schindler, U., H. Beckmann, and A.R. Cashmore, *HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region.* Plant J, 1993. **4**(1): p. 137-50.
84. Mattsson, J., et al., *A new homeobox-leucine zipper gene from Arabidopsis thaliana.* Plant Mol Biol, 1992. **18**(5): p. 1019-22.
85. Ruberti, I., et al., *A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif.* EMBO J, 1991. **10**(7): p. 1787-91.
86. Schena, M. and R.W. Davis, *HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily.* Proc Natl Acad Sci U S A, 1992. **89**(9): p. 3894-8.
87. Schena, M. and R.W. Davis, *Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families.* Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8393-7.
88. Rerie, W.G., K.A. Feldmann, and M.D. Marks, *The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis.* Genes Dev, 1994. **8**(12): p. 1388-99.
89. Reiser, L., P. Sanchez-Baracaldo, and S. Hake, *Knots in the family tree: evolutionary relationships and functions of knox homeobox genes.* Plant Mol Biol, 2000. **42**(1): p. 151-66.
90. Vollbrecht, E., et al., *The developmental gene Knotted-1 is a member of a maize homeobox gene family.* Nature, 1991. **350**(6315): p. 241-3.
91. Matsuoka, M., et al., *Expression of a rice homeobox gene causes altered morphology of transgenic plants.* Plant Cell, 1993. **5**(9): p. 1039-48.

92. Ma, H., M.D. McMullen, and J.J. Finer, *Identification of a homeobox-containing gene with enhanced expression during soybean (Glycine max L.) somatic embryo development*. Plant Mol Biol, 1994. **24**(3): p. 465-73.
93. Muller, K.J., et al., *The barley Hooded mutation caused by a duplication in a homeobox gene intron*. Nature, 1995. **374**(6524): p. 727-30.
94. Hareven, D., et al., *The making of a compound leaf: genetic manipulation of leaf architecture in tomato*. Cell, 1996. **84**(5): p. 735-44.
95. Tamaoki, M., et al., *Ectopic expression of a tobacco homeobox gene, NTH15, dramatically alters leaf morphology and hormone levels in transgenic tobacco*. Plant Cell Physiol, 1997. **38**(8): p. 917-27.
96. Guillet-Claude, C., et al., *The evolutionary implications of knox-I gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence*. Mol Biol Evol, 2004. **21**(12): p. 2232-45.
97. Burglin, T.R., *Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals*. Nucleic Acids Res, 1997. **25**(21): p. 4173-80.
98. Nagasaki, H., et al., *Functional analysis of the conserved domains of a rice KNOX homeodomain protein, OSH15*. Plant Cell, 2001. **13**(9): p. 2085-98.
99. Sakamoto, T., et al., *The conserved KNOX domain mediates specificity of tobacco KNOTTED1-type homeodomain proteins*. Plant Cell, 1999. **11**(8): p. 1419-32.
100. Bellaoui, M., et al., *The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals*. Plant Cell, 2001. **13**(11): p. 2455-70.
101. Chen, H., et al., *Interacting transcription factors from the three-amino acid loop extension superclass regulate tuber formation*. Plant Physiol, 2003. **132**(3): p. 1391-404.
102. Smith, H.M., I. Boschke, and S. Hake, *Selective interaction of plant homeodomain proteins mediates high DNA-binding affinity*. Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9579-84.
103. Smith, H.M. and S. Hake, *The interaction of two homeobox genes, BREVIPEDICELLUS and PENNYWISE, regulates internode patterning in the Arabidopsis inflorescence*. Plant Cell, 2003. **15**(8): p. 1717-27.
104. Byrne, M.E., et al., *Phyllotactic pattern and stem cell fate are determined by the Arabidopsis homeobox gene BELLRINGER*. Development, 2003. **130**(17): p. 3941-50.

105. Bhatt, A.M., et al., *VAAMANA--a BELI-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in Arabidopsis*. *Gene*, 2004. **328**: p. 103-11.
106. Mushegian, A.R. and E.V. Koonin, *Sequence analysis of eukaryotic developmental proteins: ancient and novel domains*. *Genetics*, 1996. **144**(2): p. 817-28.
107. Meisel, L. and E. Lam, *The conserved ELK-homeodomain of KNOTTED-1 contains two regions that signal nuclear localization*. *Plant Mol Biol*, 1996. **30**(1): p. 1-14.
108. Treisman, J., et al., *A single amino acid can determine the DNA binding specificity of homeodomain proteins*. *Cell*, 1989. **59**(3): p. 553-62.
109. Passner, J.M., et al., *Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex*. *Nature*, 1999. **397**(6721): p. 714-9.
110. Qian, Y.Q., et al., *The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors*. *Cell*, 1989. **59**(3): p. 573-80.
111. Abu-Shaar, M., H.D. Ryoo, and R.S. Mann, *Control of the nuclear localization of Extradenticle by competing nuclear import and export signals*. *Genes Dev*, 1999. **13**(8): p. 935-45.
112. Chang, C.P., et al., *Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins*. *Mol Cell Biol*, 1997. **17**(10): p. 5679-87.
113. Magnani, E. and S. Hake, *KNOX Lost the OX: The Arabidopsis KNATM Gene Defines a Novel Class of KNOX Transcriptional Regulators Missing the Homeodomain*. *Plant Cell*, 2008. **20**(4): p. 875-87.
114. Muller, J., et al., *In vitro interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of Knox gene function*. *Plant J*, 2001. **27**(1): p. 13-23.
115. Yang, Y., et al., *Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription*. *J Biol Chem*, 2000. **275**(27): p. 20734-41.
116. Li, T., et al., *Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA*. *Science*, 1995. **270**(5234): p. 262-9.
117. Serikawa, K.A., et al., *Localization of expression of KNAT3, a class 2 knotted1-like gene*. *Plant J*, 1997. **11**(4): p. 853-61.

118. Schmid, M., et al., *A gene expression map of Arabidopsis thaliana development*. Nat Genet, 2005. **37**(5): p. 501-6.
119. Dockx, J., et al., *The homeobox gene ATK1 of Arabidopsis thaliana is expressed in the shoot apex of the seedling and in flowers and inflorescence stems of mature plants*. Plant Mol Biol, 1995. **28**(4): p. 723-37.
120. Nishimura, A., et al., *The expression of tobacco knotted1-type class 1 homeobox genes correspond to regions predicted by the cytohistological zonation model*. Plant J, 1999. **18**(4): p. 337-47.
121. Janssen, B.J., L. Lund, and N. Sinha, *Overexpression of a homeobox gene, LeT6, reveals indeterminate features in the tomato compound leaf*. Plant Physiol, 1998. **117**(3): p. 771-86.
122. Serikawa, K.A., A. Martinez-Laborda, and P. Zambryski, *Three knotted1-like homeobox genes in Arabidopsis*. Plant Mol Biol, 1996. **32**(4): p. 673-83.
123. Tamaoki, M., et al., *Alternative RNA products from a rice homeobox gene*. Plant J, 1995. **7**(6): p. 927-38.
124. Truernit, E., et al., *A map of KNAT gene expression in the Arabidopsis root*. Plant Mol Biol, 2006. **60**(1): p. 1-20.
125. Hackbusch, J., et al., *A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins*. Proc Natl Acad Sci U S A, 2005. **102**(13): p. 4908-12.
126. Berthelsen, J., et al., *The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH*. Genes Dev, 1999. **13**(8): p. 946-53.
127. Bischof, L.J., et al., *Members of the meis1 and pbx homeodomain protein families cooperatively bind a cAMP-responsive sequence (CRS1) from bovine CYP17*. J Biol Chem, 1998. **273**(14): p. 7941-8.
128. Knoepfler, P.S., et al., *Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1*. Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14553-8.
129. Becker, A., et al., *Ancestry and diversity of BEL1-like homeobox genes revealed by gymnosperm (Gnetum gnemon) homologs*. Dev Genes Evol, 2002. **212**(9): p. 452-7.
130. Kanrar, S., O. Onguka, and H.M. Smith, *Arabidopsis inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers*. Planta, 2006. **224**(5): p. 1163-73.

131. Bao, X., et al., *Repression of AGAMOUS by BELLRINGER in floral and inflorescence meristems*. Plant Cell, 2004. **16**(6): p. 1478-89.
132. Roeder, A.H., C. Ferrandiz, and M.F. Yanofsky, *The role of the REPLUMLESS homeodomain protein in patterning the Arabidopsis fruit*. Curr Biol, 2003. **13**(18): p. 1630-5.
133. Piper, D.E., et al., *Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation*. Cell, 1999. **96**(4): p. 587-97.
134. Ryan, C.A., et al., *Polypeptide hormones*. Plant Cell, 2002. **14 Suppl**: p. S251-64.
135. Lucas, W.J., et al., *Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata*. Science, 1995. **270**(5244): p. 1980-3.
136. Sessions, A., M.F. Yanofsky, and D. Weigel, *Cell-cell signaling and movement by the floral transcription factors LEAFY and APETALA1*. Science, 2000. **289**(5480): p. 779-82.
137. Nakajima, K., et al., *Intercellular movement of the putative transcription factor SHR in root patterning*. Nature, 2001. **413**(6853): p. 307-11.
138. Kim, J.Y., et al., *Intercellular trafficking of a KNOTTED1 green fluorescent protein fusion in the leaf and shoot meristem of Arabidopsis*. Proc Natl Acad Sci U S A, 2002. **99**(6): p. 4103-8.
139. Wada, T., et al., *Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation*. Development, 2002. **129**(23): p. 5409-19.
140. Jackson, D. and J.Y. Kim, *Dispatch. Intercellular signaling: an elusive player steps forth*. Curr Biol, 2003. **13**(9): p. R349-50.
141. Ruiz-Medrano, R., B. Xoconostle-Cazares, and W.J. Lucas, *Phloem long-distance transport of CmNACP mRNA: implications for supracellular regulation in plants*. Development, 1999. **126**(20): p. 4405-19.
142. Kim, M., et al., *Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato*. Science, 2001. **293**(5528): p. 287-9.
143. Haywood, V., et al., *Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development*. Plant J, 2005. **42**(1): p. 49-68.
144. Banerjee, A.K., et al., *Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway*. Plant Cell, 2006. **18**(12): p. 3443-57.

145. Lough, T.J. and W.J. Lucas, *Integrative plant biology: role of phloem long-distance macromolecular trafficking*. *Annu Rev Plant Biol*, 2006. **57**: p. 203-32.
146. Perbal, M.C., et al., *Non-cell-autonomous function of the Antirrhinum floral homeotic proteins DEFICIENS and GLOBOSA is exerted by their polar cell-to-cell trafficking*. *Development*, 1996. **122**(11): p. 3433-41.
147. Balachandran, S., et al., *Phloem sap proteins from cucurbita maxima and ricinus communis have the capacity to traffic cell to cell through plasmodesmata*. *Proc Natl Acad Sci U S A*, 1997. **94**(25): p. 14150-5.
148. Hake, S.C. and M. Freeling, *Analysis of genetic mosaics shows that the extra epidermal cell divisions in Knotted mutant maize plants are induced by adjacent mesophyll cells*. *Nature* 1986. **320**: p. 621-623.
149. Sinha, N. and S. Hake, *Mutant characters of knotted maize leaves are determined in the innermost tissue layers*. *Dev Biol*, 1990. **141**(1): p. 203-10.
150. Kim, J.Y., Z. Yuan, and D. Jackson, *Developmental regulation and significance of KNOX protein trafficking in Arabidopsis*. *Development*, 2003. **130**(18): p. 4351-62.
151. Kim, J.Y., et al., *A novel cell-to-cell trafficking assay indicates that the KNOX homeodomain is necessary and sufficient for intercellular protein and mRNA trafficking*. *Genes Dev*, 2005. **19**(7): p. 788-93.
152. Reinhardt, D., T. Mandel, and C. Kuhlemeier, *Auxin regulates the initiation and radial position of plant lateral organs*. *Plant Cell*, 2000. **12**(4): p. 507-18.
153. Kusaba, S., et al., *Alteration of hormone levels in transgenic tobacco plants overexpressing the rice homeobox gene OSH1*. *Plant Physiol*, 1998. **116**(2): p. 471-6.
154. Scanlon, M.J., D.C. Henderson, and B. Bernstein, *SEMAPHORE1 functions during the regulation of ancestrally duplicated knox genes and polar auxin transport in maize*. *Development*, 2002. **129**(11): p. 2663-73.
155. Hooley, R., *Gibberellins: perception, transduction and responses*. *Plant Mol Biol*, 1994. **26**(5): p. 1529-55.
156. Tanaka-Ueguchi, M., et al., *Over-expression of a tobacco homeobox gene, NTH15, decreases the expression of a gibberellin biosynthetic gene encoding GA 20-oxidase*. *Plant J*, 1998. **15**(3): p. 391-400.
157. Sakamoto, T., et al., *KNOX homeodomain protein directly suppresses the expression of a gibberellin biosynthetic gene in the tobacco shoot apical meristem*. *Genes Dev*, 2001. **15**(5): p. 581-90.

158. Sakamoto, T., et al., *Ectopic expression of KNOTTED1-like homeobox protein induces expression of cytokinin biosynthesis genes in rice*. *Plant Physiol*, 2006. **142**(1): p. 54-62.
159. Hay, A., et al., *The gibberellin pathway mediates KNOTTED1-type homeobox function in plants with different body plans*. *Curr Biol*, 2002. **12**(18): p. 1557-65.
160. Xu, X., et al., *The role of gibberellin, abscisic acid, and sucrose in the regulation of potato tuber formation in vitro*. *Plant Physiol*, 1998. **117**(2): p. 575-84.
161. Rupp, H.M., et al., *Increased steady state mRNA levels of the STM and KNAT1 homeobox genes in cytokinin overproducing Arabidopsis thaliana indicate a role for cytokinins in the shoot apical meristem*. *Plant J*, 1999. **18**(5): p. 557-63.
162. Frank, M., et al., *Hormone autotrophic growth and differentiation identifies mutant lines of Arabidopsis with altered cytokinin and auxin content or signaling*. *Plant Physiol*, 2000. **122**(3): p. 721-9.
163. Hewelt, A., et al., *Ectopic expression of maize knotted1 results in the cytokinin-autotrophic growth of cultured tobacco tissues*. *Planta*, 2000. **210**(6): p. 884-9.
164. Frugis, G., et al., *Overexpression of KNAT1 in lettuce shifts leaf determinate growth to a shoot-like indeterminate growth associated with an accumulation of isopentenyl-type cytokinins*. *Plant Physiol*, 2001. **126**(4): p. 1370-80.
165. Hamant, O., et al., *The KNAT2 homeodomain protein interacts with ethylene and cytokinin signaling*. *Plant Physiol*, 2002. **130**(2): p. 657-65.
166. Zhang, J., *Evolution by gene duplication: an update*. *Trends in Ecology and Evolution*, 2003. **18**(6): p. 292-298.
167. Nei, M., I.B. Rogozin, and H. Piontkivska, *Purifying selection and birth-and-death evolution in the ubiquitin gene family*. *Proc Natl Acad Sci U S A*, 2000. **97**(20): p. 10866-71.
168. Piontkivska, H., A.P. Rooney, and M. Nei, *Purifying selection and birth-and-death evolution in the histone H4 gene family*. *Mol Biol Evol*, 2002. **19**(5): p. 689-97.
169. Zhang, J., H.F. Rosenberg, and M. Nei, *Positive Darwinian selection after gene duplication in primate ribonuclease genes*. *Proc Natl Acad Sci U S A*, 1998. **95**(7): p. 3708-13.
170. Hughes, A.L., *The evolution of functionally novel proteins after gene duplication*. *Proc Biol Sci*, 1994. **256**(1346): p. 119-24.
171. Force, A., et al., *Preservation of duplicate genes by complementary, degenerative mutations*. *Genetics*, 1999. **151**(4): p. 1531-45.

172. Lynch, M., *Genomics. Gene duplication and evolution*. Science, 2002. **297**(5583): p. 945-7.
173. Long, M., *Evolution of novel genes*. Curr Opin Genet Dev, 2001. **11**(6): p. 673-80.
174. Drouin, G. and G.A. Dover, *Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes*. J Mol Evol, 1990. **31**(2): p. 132-50.
175. Loguercio, L.L. and T.A. Wilkins, *Structural analysis of a hmg-coA-reductase pseudogene: insights into evolutionary processes affecting the hmgr gene family in allotetraploid cotton (Gossypium hirsutum L.)*. Curr Genet, 1998. **34**(4): p. 241-9.
176. Mladek, C., K. Guger, and M.T. Hauser, *Identification and characterization of the ARIADNE gene family in Arabidopsis. A group of putative E3 ligases*. Plant Physiol, 2003. **131**(1): p. 27-40.
177. Kvarnheden, A., K. Tandre, and P. Engstrom, *A cdc2 homologue and closely related processed retropseudogenes from Norway spruce*. Plant Mol Biol, 1995. **27**(2): p. 391-403.
178. Benovoy, D. and G. Drouin, *Processed pseudogenes, processed genes, and spontaneous mutations in the Arabidopsis genome*. J Mol Evol, 2006. **62**(5): p. 511-22.
179. Champagne, C.E.M. and N.W. Ashton, *Ancestry of KNOX genes revealed by bryophyte (Physcomitrella patens) homologs*. New phytologist, 2001. **150**: p. 23-36.
180. Bowe, L.M., G. Coat, and C.W. dePamphilis, *Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers*. Proc Natl Acad Sci U S A, 2000. **97**(8): p. 4092-7.
181. Stuessy, T.F., *A transitional-combinational theory for the origin of angiosperms*. Taxon, 2004. **53**: p. 3-16.
182. Miller, C.N., ed. *The origin of modern conifer families*. Origin and evolution of Gymnosperms., ed. C.B. Beck. 1988, Columbia University Press, New York. 448-486.
183. Kinlaw, C.N., DB., *Complex gene families in pine genomes*. Trends in Plant Science, 1997. **2**: p. 356-359.
184. Doebley, J. and L. Lukens, *Transcriptional regulators and the evolution of plant form*. Plant Cell, 1998. **10**(7): p. 1075-82.

185. Doebley, J., A. Stec, and L. Hubbard, *The evolution of apical dominance in maize*. Nature, 1997. **386**(6624): p. 485-8.
186. Frary, A., et al., *fw2.2: a quantitative trait locus key to the evolution of tomato fruit size*. Science, 2000. **289**(5476): p. 85-8.
187. Li, C., A. Zhou, and T. Sang, *Rice domestication by reducing shattering*. Science, 2006. **311**(5769): p. 1936-9.
188. Doebley, J.F., B.S. Gaut, and B.D. Smith, *The molecular genetics of crop domestication*. Cell, 2006. **127**(7): p. 1309-21.
189. Cairney, J.P., GS., *The cellular and molecular biology of conifer embryogenesis*. New phytologist, 2007. **176**: p. 511-536.
190. Lindsey, K. and J. Topping, *Embryogenesis: a Question of Pattern*. Journal of Experimental Botany, 1993. **44**(2): p. 359-374.
191. Ingouff, M., et al., *The molecular characterization of PaHB2, a homeobox gene of the HD-GL2 family expressed during embryo development in Norway spruce*. J Exp Bot, 2003. **54**(386): p. 1343-50.
192. Oh, T.J., et al., *Evidence for stage-specific modulation of specific microRNAs (miRNAs) and miRNA processing components in zygotic embryo and female gametophyte of loblolly pine (Pinus taeda)*. New Phytol, 2008. **179**(1): p. 67-80.
193. Sato, Y., Y. Fukuda, and H.Y. Hirano, *Mutations that cause amino acid substitutions at the invariant positions in homeodomain of OSH3 KNOX protein suggest artificial selection during rice domestication*. Genes Genet Syst, 2001. **76**(6): p. 381-92.
194. Grotkopp, E., et al., *Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analyses*. Evolution Int J Org Evolution, 2004. **58**(8): p. 1705-29.
195. Zhang, H., J.Y. Lee, and B. Tian, *Biased alternative polyadenylation in human tissues*. Genome Biol, 2005. **6**(12): p. R100.
196. Tian, B., Z. Pan, and J.Y. Lee, *Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing*. Genome Res, 2007. **17**(2): p. 156-65.
197. Wachter, A., et al., *Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs*. Plant Cell, 2007. **19**(11): p. 3437-50.
198. Greene, B., R. Walko, and S. Hake, *Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations*. Genetics, 1994. **138**(4): p. 1275-85.

199. Luehrsen, K.R. and V. Walbot, *Insertion of Mu1 elements in the first intron of the Adh1-S gene of maize results in novel RNA processing events*. *Plant Cell*, 1990. **2**(12): p. 1225-38.
200. Luehrsen, K.R. and V. Walbot, *Insertion of non-intron sequence into maize introns interferes with splicing*. *Nucleic Acids Res*, 1992. **20**(19): p. 5181-7.
201. Mann, R.S. and G. Morata, *The developmental and molecular biology of genes that subdivide the body of Drosophila*. *Annu Rev Cell Dev Biol*, 2000. **16**: p. 243-71.
202. Zhang, Y. and M.A. Frohman, *Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs*. *Methods Mol Biol*, 1997. **69**: p. 61-87.
203. Bulyk, M.L., *Computational prediction of transcription-factor binding site locations*. *Genome Biol*, 2003. **5**(1): p. 201.
204. Wasserman, W.W., et al., *Human-mouse genome comparisons to locate regulatory sites*. *Nat Genet*, 2000. **26**(2): p. 225-8.
205. Weitzman, J.B., *Tracking evolution's footprints in the genome*. *J Biol*, 2003. **2**(2): p. 9.
206. Zhang, Z. and M. Gerstein, *Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements*. *J Biol*, 2003. **2**(2): p. 11.
207. Hay, A. and M. Tsiantis, *The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta*. *Nat Genet*, 2006. **38**(8): p. 942-7.
208. Inada, D.C., et al., *Conserved noncoding sequences in the grasses*. *Genome Res*, 2003. **13**(9): p. 2030-41.
209. Morimoto, R., et al., *Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene Wknox1 in common wheat*. *Plant Mol Biol*, 2005. **57**(6): p. 907-24.
210. Blanchette, M. and M. Tompa, *FootPrinter: A program designed for phylogenetic footprinting*. *Nucleic Acids Res*, 2003. **31**(13): p. 3840-2.
211. Mahony, S., P.E. Auron, and P.V. Benos, *DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies*. *PLoS Comput Biol*, 2007a. **3**(3): p. e61.
212. Mahony, S. and P.V. Benos, *STAMP: a web tool for exploring DNA-binding motif similarities*. *Nucleic Acids Res*, 2007b. **35**(Web Server issue): p. W253-8.

213. Rombauts, S., et al., *Computational approaches to identify promoters and cis-regulatory elements in plant genomes*. *Plant Physiol*, 2003. **132**(3): p. 1162-76.
214. Higo, K., et al., *Plant cis-acting regulatory DNA elements (PLACE) database: 1999*. *Nucleic Acids Res*, 1999. **27**(1): p. 297-300.
215. Kawaji, H., et al., *Dynamic usage of transcription start sites within core promoters*. *Genome Biol*, 2006. **7**(12): p. R118.
216. Roni, V., R. Carpio, and B. Wissinger, *Mapping of transcription start sites of human retina expressed genes*. *BMC Genomics*, 2007. **8**: p. 42.
217. Roberts, S., T. Purton, and D.L. Bentley, *A protein-binding site in the c-myc promoter functions as a terminator of RNA polymerase II transcription*. *Genes Dev*, 1992. **6**(8): p. 1562-74.
218. Winter, J., et al., *Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1*. *BMC Mol Biol*, 2007. **8**: p. 105.
219. Liu, H. and E.M. Johnson, *Distinct proteins encoded by alternative transcripts of the PURG gene, located contrapodal to WRN on chromosome 8, determined by differential termination/polyadenylation*. *Nucleic Acids Res*, 2002. **30**(11): p. 2417-26.
220. Kiss, T., *Biogenesis of small nuclear RNPs*. *J Cell Sci*, 2004. **117**(Pt 25): p. 5949-51.
221. Do, C.B., D.A. Woods, and S. Batzoglou, *CONTRAFold: RNA secondary structure prediction without physics-based models*. *Bioinformatics*, 2006. **22**(14): p. e90-8.
222. Luo, H., et al., *Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses*. *Plant Biol (Stuttg)*, 2005. **7**(5): p. 459-68.
223. Nakashima, K., et al., *Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis*. *Plant Mol Biol*, 2006. **60**(1): p. 51-68.
224. Abe, H., et al., *Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression*. *Plant Cell*, 1997. **9**(10): p. 1859-68.
225. Byrne, M.E., et al., *Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis*. *Nature*, 2000. **408**(6815): p. 967-71.

226. Hudson, M.E., D.R. Lisch, and P.H. Quail, *The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway*. Plant J, 2003. **34**(4): p. 453-71.
227. Werneke, J.M., J.M. Chatfield, and W.L. Ogren, *Alternative mRNA splicing generates the two ribulosebiphosphate carboxylase/oxygenase activase polypeptides in spinach and Arabidopsis*. Plant Cell, 1989. **1**(8): p. 815-25.
228. Lam, E., et al., *A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter*. Plant Cell, 1990. **2**(9): p. 857-66.
229. Washio, K., *Functional dissections between GAMYB and Dof transcription factors suggest a role for protein-protein associations in the gibberellin-mediated expression of the RAmylA gene in the rice aleurone*. Plant Physiol, 2003. **133**(2): p. 850-63.
230. Morita, A., et al., *Functional dissection of a sugar-repressed alpha-amylase gene (RAmyl A) promoter in rice embryos*. FEBS Lett, 1998. **423**(1): p. 81-5.
231. Huang, N., et al., *Classification and characterization of the rice alpha-amylase multigene family*. Plant Mol Biol, 1990. **14**(5): p. 655-68.
232. Liu, F., et al., *Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis*. Plant Physiol, 2005. **137**(3): p. 1115-29.
233. Olive, M.R., et al., *Functional properties of the anaerobic responsive element of the maize Adh1 gene*. Plant Mol Biol, 1990. **15**(4): p. 593-604.
234. Olive, M.R., W.J. Peacock, and E.S. Dennis, *The anaerobic responsive element contains two GC-rich sequences essential for binding a nuclear protein and hypoxic activation of the maize Adh1 promoter*. Nucleic Acids Res, 1991. **19**(25): p. 7053-60.
235. Perry, D.J. and G.R. Furnier, *Pinus banksiana has at least seven expressed alcohol dehydrogenase genes in two linked groups*. Proc Natl Acad Sci U S A, 1996. **93**(23): p. 13020-3.
236. Agarwal, M., et al., *Arabidopsis thaliana Hsp100 proteins: kith and kin*. Cell Stress Chaperones, 2001. **6**(3): p. 219-24.
237. Batra, G., et al., *Complexity of rice Hsp100 gene family: lessons from rice genome sequence data*. J Biosci, 2007. **32**(3): p. 611-9.
238. Danpure, C.J., *How can the products of a single gene be localized to more than one intracellular compartment?* Trends Cell Biol, 1995. **5**(6): p. 230-8.

239. Quesada, V., et al., *Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time*. EMBO J, 2003. **22**(12): p. 3142-52.
240. Gromak, N., S. West, and N.J. Proudfoot, *Pause sites promote transcriptional termination of mammalian RNA polymerase II*. Mol Cell Biol, 2006. **26**(10): p. 3986-96.
241. Mignone, F., et al., *UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs*. Nucleic Acids Res, 2005. **33**(Database issue): p. D141-6.
242. Markel, H., J. Chandler, and W. Werr, *Translational fusions with the engrailed repressor domain efficiently convert plant transcription factors into dominant-negative functions*. Nucleic Acids Res, 2002. **30**(21): p. 4709-19.
243. An, G., M.A. Costa, and S.B. Ha, *Nopaline synthase promoter is wound inducible and auxin inducible*. Plant Cell, 1990. **2**(3): p. 225-33.
244. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana*. Plant J, 1998. **16**(6): p. 735-43.
245. Bechtold, N., J. Ellis, and G. Pelletier, *In planta Agrobacterium-mediated gene transfer by infiltration of adult Arabidopsis thaliana plants*. C.R. Acad. Sci., 1993. **316**: p. 1194-1199.
246. Haseloff. [cited 2008; Available from: http://www.plantsci.cam.ac.uk/Haseloff/teaching/PlantSciPartII_2006/LectureIntro/Lecture1/files/page1-1068-full.html].
247. Tang, W., *Agrobacterium-mediated transformation and assessment of factors influencing transgene expression in loblolly pine (Pinus taeda L.)*. Cell Res, 2001. **11**(3): p. 237-43.
248. Tang, W. and Y. Tian, *Transgenic loblolly pine (Pinus taeda L.) plants expressing a modified delta-endotoxin gene of Bacillus thuringiensis with enhanced resistance to Dendrolimus punctatus Walker and Crypyothelea formosicola Staud.* J Exp Bot, 2003. **54**(383): p. 835-44.
249. Wielopolska, A., et al., *A high-throughput inducible RNAi vector for plants*. Plant Biotechnol J, 2005. **3**(6): p. 583-90.
250. Clapham, D., et al., *Gene Transfer by Particle Bombardment to Embryogenic Cultures of Picea abies and the Production of Transgenic Plantlets*. Scandinavian Journal of Forest Research, 2000. **15**(2).
251. Kimura, S., et al., *Natural variation in leaf morphology results from mutation of a novel KNOX gene*. Curr Biol, 2008. **18**(9): p. 672-7.

252. McManaway, M.E., et al., *Tumour-specific inhibition of lymphoma growth by an antisense oligodeoxynucleotide*. *Lancet*, 1990. **335**(8693): p. 808-11.
253. Kulka, M., et al., *Site specificity of the inhibitory effects of oligo(nucleoside methylphosphonate)s complementary to the acceptor splice junction of herpes simplex virus type 1 immediate early mRNA 4*. *Proc Natl Acad Sci U S A*, 1989. **86**(18): p. 6868-72.
254. Gniadkowski, M., et al., *Characterization of intronic uridine-rich sequence elements acting as possible targets for nuclear proteins during pre-mRNA splicing in Nicotiana plumbaginifolia*. *Nucleic Acids Res*, 1996. **24**(4): p. 619-27.
255. Ciavatta, V.T., et al., *A promoter from the loblolly pine PtNIP1;1 gene directs expression in an early-embryogenesis and suspensor-specific fashion*. *Planta*, 2002. **215**(4): p. 694-8.